

Low-rank Optimization with Convex Constraints

CHRISTIAN GRUSSLER *, ANDERS RANTZER *, AND PONTUS GISELSSON *

Abstract. The problem of low-rank approximation with convex constraints, which often appears in data analysis, image compression and model order reduction, is considered. Given a data matrix, the objective is to find an approximation of desired lower rank that fulfills the convex constraints and minimizes the distance to the data matrix in the Frobenius norm. The problem of matrix completion can be seen as a special case of this.

Today, one of the most widely used techniques is to approximate this non-convex problem using convex nuclear norm regularization. In many situations, this technique does not give solutions with desirable properties. We instead propose to use the largest convex minorizer (under-approximation) of the Frobenius norm and the rank constraint as a convex proxy. This optimal convex proxy can be combined with other convex constraints to form an optimal convex minorizer of the original non-convex problem. With this approach, we get easily verifiable conditions under which the solutions to the convex relaxation and the original non-convex problem coincide. Several numerical examples are provided for which that is the case. We also see that our proposed convex relaxation consistently performs better than the nuclear norm heuristic, especially in the matrix completion case.

The expressibility and computational tractability is of great importance for a convex relaxation. We provide a closed-form expression for the proposed convex approximation, and show how to represent it as a semi-definite program. We also show how to compute the proximal operator of the convex approximation. This allows us to use scalable first-order methods to solve convex approximation problems of large size.

Key words. Low-rank Approximation, Douglas-Rachford Splitting, Matrix Completion, k -support norm, Semi-Definite Programming, Non-Convex Optimization.

AMS subject classifications. 90C06, 90C25, 90C26, 90C59

1. Introduction. The main reason for low-rank approximation theory lies in the idea of studying only the few essential components of an otherwise complex operator. For instance, it is well-known that the rank of a matrix $N \in \mathbb{R}^{n \times m}$ equals the dimension of its column space. In other words, if a matrix has low rank then only a small number of basis vectors is needed to span its range and a possibly high dimensional subspace in \mathbb{R}^m can be disregarded when studying $y = Nx$. Hence, if N is sufficiently close to a lower rank matrix, it may be sufficient to study the approximation $y \approx \hat{N}x$ where $\text{rank}(\hat{N}) < \text{rank}(N)$.

With this concept in mind, one can understand why many areas such as image analysis, model order reduction, multivariate linear regression, etc. desire a low-rank approximation (see [3, 12, 11, 14, 33, 40, 49, 44, 43, 52, 56, 64, 65]). In Sections 6 to 9 some of these applications are explained in greater depth.

For unitarily invariant norms an optimal low-rank approximation can be found by performing a singular value decomposition (SVD) (see Section 2). Unfortunately, these approximations usually do not fulfill structural constraints such as element-wise non-negativity, Hankel-structure, prescribed entries, etc. (see [7, 12, 16, 36, 49, 52, 56, 64]). Only for a few known cases an explicit solution to the constrained low-rank approximation problem can be determined (see [3, 49, 64]). To this end, other concepts based on convex optimization have been developed (see [14, 24, 44, 43, 56]). Many of them rely on nuclear norm regularization, which allows to incorporate any convex constraint (see Subsection 5.1). Nevertheless, the question of optimality is not addressed, unless one aims for a minimum rank solution (see [12, 56]). Besides the nuclear norm heuristic, other commonly used heuristics, e.g. for element-wise non-negativity are briefly considered in Section 6.

In this work, we study the optimal low-rank approximation problem with a prescribed target rank and convex constraints (see Problem 1). This is a continuation of the authors work [31]. It is shown that a globally optimal solution to our non-convex problem can often be

*Department of Automatic Control, Lund University, Sweden (christiang@control.lth.se, rantzer@control.lth.se, pontusg@control.lth.se)

determined by convex optimization (see Section 3). In particular, if the SVD-approximation of a matrix is unique, then it is a solution to a semi-definite program (SDP). Even though the approach presented can be linked to the regularization method in [44, 43], we will see that it is essentially parameter free.

In Section 4 some computational aspects of the convexified problem are discussed. First, an SDP-representation of the convex proxy is presented, which allows us to compute solutions for small scale examples with SDP-representable constraints. Subsequently, we derive the so-called Douglas-Rachford iterations in order to deal with examples of larger size and sufficiently simple constraints (see Subsection 4.2). As a consequence, we will be able to prove local convergence of the Douglas-Rachford iterations of the originally non-convex problem.

The paper is organized as follows. In Section 2 we recap the unconstrained low-rank approximation problem and define our main problem. The main approach is derived and discussed in Section 3 with some computational aspects examined in Section 4. Other known approaches, including the nuclear norm heuristic are discussed in Section 5. In Sections 6 to 9 some applications are presented that show the usefulness of this approach. Moreover, the examples are chosen to illustrate some properties and drawbacks of this method. Finally, we draw a conclusion and discuss future research in Section 10.

2. Background. The following notation for real matrices and vectors $X = (x_{ij}) \in \mathbb{R}^{n \times m}$ is used throughout this paper. If $X = X^T$, then we write $X \in \mathbb{S}$. Moreover, if X is positive definite (semi-definite) we write $X \succ 0$ ($X \succeq 0$). We also use these notations to describe the relation between two matrices, e.g. $A \succeq B$ means $A - B \succeq 0$. The non-increasingly ordered singular values of $X \in \mathbb{R}^{n \times m}$, counted with multiplicity, are denoted by $\sigma_1(X) \geq \dots \geq \sigma_{\min\{m,n\}}(X)$. Further we define by $\langle X, Y \rangle := \sum_{i=1}^m \sum_{j=1}^n x_{ij}y_{ij} = \text{trace}(X^T Y)$, $X, Y \in \mathbb{R}^{n \times m}$ the Frobenius inner-product on the Hilbert space $\mathbb{R}^{n \times m}$. Correspondingly, the Frobenius norm is defined as

$$\|X\|_F := \sqrt{\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2(X)}.$$

The Frobenius norm is unitarily invariant, i.e. $\|UXV\|_F = \|X\|_F$ for all unitary matrices U and V . A complete characterization of all unitarily invariant norms is given in [39].

This work mainly considers the norms that are found in the following Lemma, which is proven in Subsection A.3.

LEMMA 1. *Let $M \in \mathbb{R}^{n \times m}$, $r \in \mathbb{N}$ such that $1 \leq r \leq q := \min\{m, n\}$ and \mathcal{P}_r denote the set of all orthogonal projections of rank r . Then*

$$(1) \quad \|M\|_r := \sqrt{\sum_{i=1}^r \sigma_i^2(M)} = \sqrt{\max_{P \in \mathcal{P}_r} \langle P, M^T M \rangle}$$

is a unitarily invariant norm with dual norm

$$\|M\|_{r*} := \max_{\|X\|_r \leq 1} \langle M, X \rangle = \max_{\sum_i^r s_i^2 \leq 1} \left[\sum_{i=1}^r \sigma_i(M) s_i + s_r \sum_{i=r+1}^q \sigma_i(M) \right].$$

Moreover,

$$(2) \quad \|M\|_1 \leq \dots \leq \|M\|_q = \|M\|_F = \|M\|_{q*} \leq \dots \leq \|M\|_{1*}.$$

$$(3) \quad \text{rank}(M) \leq r \text{ if and only if } \|M\|_r = \|M\|_F = \|M\|_{*r}.$$

An explicit expression of $\|\cdot\|_{r*}$ was first derived in [25]. Notice that $\|M\|_1 = \sigma_1(M)$ is equal to the spectral norm, and its dual norm $\|M\|_{1*} = \sum_{i=1}^{\min\{m,n\}} \sigma_i(M)$ is equal to the nuclear/trace norm. These norms can be formulated using convex linear matrix inequalities (see [24, 56]). In Section 3 we will see that the same holds true for $\|\cdot\|_r$ and their duals.

Recently, the vector version of the $r*$ norm has appeared as "k-support norm" (see [4]). As a result, some authors have adopted that name for the matrix case (see [22, 42, 50]). However, as for other vector/matrix norm pairings e.g. the ℓ_1 norm of the singular values is called the nuclear norm, we have chosen the $r*$ norm notation to distinguish between the matrix and vector case.

2.1. Statements. Let us turn to the underlying problem of this work. We start with the traditional optimal low-rank approximation problem in $\mathbb{R}^{n \times m}$, which is formulated as follows. Given $N \in \mathbb{R}^{n \times m}$ and $r \in \mathbb{N}$ such that $1 \leq r \leq \min\{m, n\}$, find a solution $M^* \in \mathbb{R}^{n \times m}$ to

$$(4) \quad \begin{aligned} & \text{minimize} && \|N - M\|_F^2 \\ & \text{subject to} && \text{rank}(M) \leq r \end{aligned}$$

In case of the Hilbert-Schmidt norm, the natural operator generalization of the Frobenius norm, this problem has been solved by Schmidt and generalized by Mirsky to unitarily invariant norms (see [3]). The result is stated next.

PROPOSITION 1. *Let $N \in \mathbb{R}^{n \times m}$ and $r \in \mathbb{N}$ such that $1 \leq r \leq \min\{m, n\}$, then*

$$\min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \|N - M\| = \|\text{diag}(\sigma_{r+1}(N), \dots, \sigma_{\min\{m,n\}}(N))\|,$$

holds for any unitarily invariant norm $\|\cdot\|$.

If an SVD of N is given by $N = \sum_{i=1}^{\min\{m,n\}} \sigma_i u_i v_i^T$, a solution to (4) can be derived as $M^* = \text{svd}_r(N) := \sum_{i=1}^r \sigma_i u_i v_i^T$, which we refer to as a *standard SVD-approximation*. This solution may not be unique if the norm does not depend on all singular values or if $\sigma_r(N) = \sigma_{r+1}(N)$. Nevertheless, with the Frobenius norm and $\sigma_r(N) \neq \sigma_{r+1}(N)$ the uniqueness of M^* is guaranteed, which on the other hand does not account for additional constraints.

In this work we look at the following extension of (4).

PROBLEM 1. *Given $N \in \mathbb{R}^{n \times m}$, find $M^* \in \mathbb{R}^{n \times m}$ with $\text{rank}(M^*) \leq r$ such that*

$$\min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\frac{1}{2} \|N - M\|_F^2 + g(M) \right] = \frac{1}{2} \|N - M^*\|_F^2 + g(M^*),$$

where $g : \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \cup \{\infty\}$ is a given closed proper convex functional (see Definition A.2).

Compared to (4), Problem 1 has an additional functional g that can be used to add information about the desired solution. Both problems are non-convex due to the rank constraint. Nevertheless, we will see in Section 3 that they can often be solved by convex optimization. In particular, if (4) has a unique solution, it is possible to determine it by solving a semi-definite program instead of an SVD.

In the following we often think of g as $g(M) \equiv \chi_{\mathcal{C}}(M)$ where

$$\chi_{\mathcal{C}}(M) := \begin{cases} 0, & M \in \mathcal{C} \\ \infty, & M \notin \mathcal{C} \end{cases}$$

is defined to be the characteristic function of a (convex) set $\mathcal{C} \subset \mathbb{R}^{n \times m}$ – within the optimization literature often called indicator function. We also use $\chi_{\text{rank}(M) \leq r}(M)$ to denote the characteristic function of the set of matrices which have at most rank r .

Notice that Problem 1 also deals with cases where $N = 0$, which then covers the class of matrix completion problems (see Section 7).

3. The r^* approach. In the following we consider an attempt of finding an *optimal* solution to Problem 1. It is a continuation of the authors work [31]. The insights obtained here will allow later to generalize and improve upon current standard approaches (see Section 7). The main idea is to derive a convex minorizer (under-approximation) of the non-convex cost-function in Problem 1 by means of Fenchel-duality (see Subsection A.2). We denote by f^* and f^{**} the conjugate and bi-conjugate functionals of $f : \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \cup \{\infty\}$ (see Definition A.1).

THEOREM 1. *Let $N \in \mathbb{R}^{n \times m}$ and $r \in \mathbb{N}$ such that $1 \leq r \leq \min\{m, n\}$. Then the conjugate and bi-conjugate functionals of $f(M) := \frac{1}{2} \|N - M\|_F^2 + \chi_{\text{rank}(M) \leq r}(M)$ are given by*

$$(5) \quad f^*(D) = \frac{1}{2} \|N + D\|_r^2 - \frac{1}{2} \|N\|_F^2,$$

$$(6) \quad f^{**}(M) = \frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2.$$

Proof. Let $N \in \mathbb{R}^{n \times m}$ and $f(M) := \frac{1}{2} \|N - M\|_F^2 + \chi_{\text{rank}(M) \leq r}(M)$. Then,

$$\begin{aligned} f^*(D) &= \sup_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\langle D, M \rangle - \frac{1}{2} \|N - M\|_F^2 \right] \\ &= \sup_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} -\frac{1}{2} \|N - M + D\|_F^2 + \langle D, N \rangle + \frac{1}{2} \|D\|_F^2 \\ &= -\frac{1}{2} \|N + D\|_F^2 + \frac{1}{2} \|N + D\|_r^2 + \langle D, N \rangle + \frac{1}{2} \|D\|_F^2 \\ &= -\frac{1}{2} \|N\|_F^2 + \frac{1}{2} \|N + D\|_r^2 \end{aligned}$$

where the third equality follows by Proposition 1. Hence,

$$\begin{aligned} f^{**}(M) &= \sup_{D \in \mathbb{R}^{n \times m}} \left[\langle D, M \rangle + \frac{1}{2} \|N\|_F^2 - \frac{1}{2} \|N + D\|_r^2 \right] \\ &= \sup_{D \in \mathbb{R}^{n \times m}} \left[\langle D - N, M \rangle + \frac{1}{2} \|N\|_F^2 - \frac{1}{2} \|D\|_r^2 \right] \\ &= \frac{1}{2} \|N\|_F^2 - \langle N, M \rangle + \frac{1}{2} \|M\|_{r^*}^2. \quad \square \end{aligned}$$

It is possible to show that $f(M) \geq f^{**}(M)$ for all $M \in \mathbb{R}^{n \times m}$, i.e. f^{**} is a convex minorizer of f . In fact, f^{**} is the largest convex minorizer of f (see [37, Theorem 1.3.5]), i.e. the point-wise supremum of all affine functions majorized by f (see Figure 1). This allows us to

construct the following dual and bi-dual problem to Problem 1:

$$(A) \quad - \min_{D \in \mathbb{R}^{n \times m}} \left[g^*(-D) + \frac{1}{2} \|N + D\|_r^2 - \frac{1}{2} \|N\|_F^2 \right],$$

$$(B) \quad \min_{M \in \mathbb{R}^{n \times m}} \left[\frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2 + g(M) \right].$$

Observe that $f^{**} + g$ is the largest convex minorizer of $f + g$ with g is a summand. Therefore,

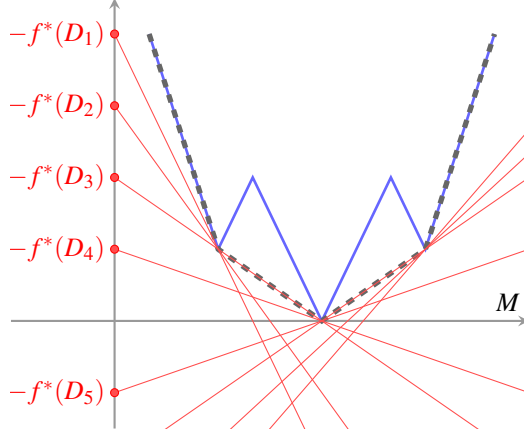


Figure 1: Schematic plot of $\text{—} f(M)$, $\text{---} f^{**}(M)$ and $\text{—} \text{tangents through } -f^*(D_i)$.

we propose to use (B) instead of the nuclear norm heuristic (see (25) in Subsection 5.1) as a convex proxy to Problem 1. We will see that it has many interesting properties and that sometimes it can be guaranteed to solve the original non-convex problem. Theorem 1 gives the following duality result through Fenchel-duality (see Lemma A.1 and Proposition A.3).

PROPOSITION 2. *Let $N \in \mathbb{R}^{n \times m}$ and $g : \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \cup \{\infty\}$ be a closed proper convex functional, then for all $r \in \mathbb{N}$ such that $1 \leq r \leq \min\{m, n\}$ one has*

$$(C) \quad \min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\frac{1}{2} \|N - M\|_F^2 + g(M) \right] \geq - \min_{D \in \mathbb{R}^{n \times m}} \left[g^*(-D) + \frac{1}{2} \|N + D\|_r^2 - \frac{1}{2} \|N\|_F^2 \right]$$

$$= \min_{M \in \mathbb{R}^{n \times m}} \left[\frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2 + g(M) \right].$$

Since the original Problem 1 is non-convex, there is a duality-gap for some choices of g (see Section 8). This is reflected by the inequality in (C). However, there are situations without a duality-gap. Next, we present a number of important cases.

In the following, the set of minimizers of a functional f over a given set S is denoted by $\text{argmin}_S f$. If $\text{argmin}_S f = \{x^*\}$ is just a singleton, we write $x^* = \text{argmin}_S f$.

PROPOSITION 3. *Assume that (B) has a minimizer M^* with $\text{rank}(M^*) \leq r$. Then,*

$$\text{argmin}_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\frac{1}{2} \|N - M\|_F^2 + g(M) \right] = \text{argmin}_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2 + g(M) \right].$$

Proof. The result follows by combining Proposition 2 with (3) in Lemma 1. \square

Thus obtaining a rank- r solution to the convex relaxation problem (B) implies solving the original non-convex problem. Another way to state this result, that gives additional insight on the solution to Problem 1, is as follows.

THEOREM 2. *Assume that D^* is a solution to (A) and $\sigma_r(N + D^*) \neq \sigma_{r+1}(N + D^*)$ or $\sigma_r(N + D^*) = 0$. Then there is no duality gap in (C) and $\text{svd}_r(N + D^*)$ is the unique minimizing argument of Problem 1, i.e.*

$$\text{svd}_r(N + D^*) = \underset{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}}{\text{argmin}} \left[\frac{1}{2} \|N - M\|_F^2 + g(M) \right].$$

This result provides a simple sufficient condition for the uniqueness of a solution to Problem 1. However, it is not a necessary condition. A proof of Theorem 2 is given in a more general setting in Theorem 3, which also allows us to say something about the rank of the convex relaxation solution if there is a duality-gap.

THEOREM 3. *Let D^* and M^* be solutions to (A) and (B), respectively. Further suppose that an SVD of $N + D^*$ is given by $N + D^* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(N + D^*) u_i v_i^T$ with $\sigma_{r-t}(N + D^*) \neq \sigma_{r-t+1}(N + D^*) = \dots = \sigma_r(N + D^*) = \dots = \sigma_{r+s}(N + D^*) \neq \sigma_{r+s+1}(N + D^*)$, where $t = r$ and $s = \min\{m,n\} - r$ if $\sigma_1 = \sigma_r$ and $\sigma_{\min\{m,n\}} = \sigma_r$, respectively. Then there exists $T \in \mathbb{R}^{s+t \times s+t}$ with $T \succeq 0$, $\|T\|_1 \leq 1$ and $\|T\|_{1*} = t$ such that*

$$M^* = \sum_{i=1}^{r-t} \sigma_i(N + D^*) u_i v_i^T + \sigma_r(N + D^*) (u_{r-t+1} \dots u_{r+s}) T (v_{r-t+1} \dots v_{r+s})^T.$$

In particular, $\text{rank}(M^) \leq r + s$ and if $\sigma_r(N + D^*) \neq \sigma_{r+1}(N + D^*)$ or $\sigma_r(N + D^*) = 0$ then $M^* = \text{svd}_r(N + D^*)$.*

Proof. If D^* and M^* are solutions to (A) and (B), respectively, then by Proposition A.3 it holds that $f^{**}(M^*) = \langle D^*, M^* \rangle - f^*(D^*)$, where f^* and f^{**} are given by (5) and (6). Hence, by Proposition A.4 it follows that

$$M^* \in \partial_D \frac{1}{2} \|N + D\|_r^2 \Big|_{D=D^*} = \|N + D^*\|_r \partial_D \|N + D\|_r \Big|_{D=D^*}$$

and invoking Proposition A.5 proves the result. \square

Observe that whenever (B) does not have a unique solution, it must hold by Proposition 1 and Theorem 3 that $\sigma_r(N + D^*) = \sigma_{r+1}(N + D^*)$ for all solutions D^* to (A). Furthermore, Theorem 3 implies that $\text{svd}_r(N)$ with $\sigma_r(N) \neq \sigma_{r+1}(N)$, can be determined by solving a convex problem.

COROLLARY 1. *Let $N \in \mathbb{R}^{n \times m}$ and $r \in \mathbb{N}$ be such that $1 \leq r \leq \min\{m,n\}$. Then*

$$\min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \frac{1}{2} \|N - M\|_F^2 = \frac{1}{2} \|N\|_F^2 - \frac{1}{2} \|N\|_r^2 = \min_{M \in \mathbb{R}^{n \times m}} \left[\frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2 \right]$$

and $\text{svd}_r(N) \in \text{argmin}_{M \in \mathbb{R}^{n \times m}} \left[\frac{1}{2} \|M\|_{r^}^2 - \langle N, M \rangle \right]$. If $\sigma_r(N) \neq \sigma_{r+1}(N)$ or $\sigma_r = 0$ then*

$$\text{svd}_r(N) = \underset{M \in \mathbb{R}^{n \times m}}{\text{argmin}} \left[\frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle \right].$$

Proof. Since $g = 0$ implies $g^*(D) < \infty \Leftrightarrow D = 0$, the result follows by Theorem 3. \square

Finally, notice that several extensions of Problem 1 are covered by the preceding results. For instance, one can consider the weighted case

$$(7) \quad \min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\frac{1}{2} \|W(N - M)\|_F^2 + g(M) \right]$$

where $W \in \mathbb{R}^{l \times n}$ and $\text{rank}(W) = n$. Let $\tilde{g}(\tilde{M}) := g(W^\dagger \tilde{M})$, where W^\dagger denotes the pseudo-inverse of W (see [39]). Since $\text{rank}(\tilde{M}) = \text{rank}(W^\dagger \tilde{M}) = \text{rank}(M)$, one can reformulate (7) such that it fits the formulation of Problem 1:

$$\min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\frac{1}{2} \|W(N - M)\|_F^2 + g(M) \right] = \min_{\substack{\tilde{M} \in \mathbb{R}^{n \times m} \\ \text{rank}(\tilde{M}) \leq r}} \left[\frac{1}{2} \|WN - \tilde{M}\|_F^2 + \tilde{g}(\tilde{M}) \right].$$

Note that $\|W(N - M)\|_F^2 = \text{trace}((N - M)^T W^T W (N - M)) =: \langle N - M, N - M \rangle_{W^T W}$ defines another inner product and norm. A suitable W may enable us to satisfy the requirements of Theorem 2, where the Frobenius norm fails to do so.

3.1. Geometric interpretation. Assuming that $g(M) \equiv \chi_{\mathcal{C}}(M)$ for some closed convex set $\mathcal{C} \subset \mathbb{R}^{n \times m}$, the preceding results offer an insightful geometric interpretation. Note that (B) has the same solutions as

$$(8) \quad \min_{\substack{M \in \mathcal{C} \\ \langle N, M \rangle = c}} \|M\|_{r*},$$

where $c := \langle N, M^* \rangle$ and M^* is a solution to (B). The solutions of (8) can be found by studying the set $B_{r*}^{\bar{\varepsilon}} \cap H \cap \mathcal{C}$ where

$$\begin{aligned} B_{r*}^{\varepsilon} &:= \{X : \|X\|_{r*} \leq \varepsilon\}, \\ H &:= \{X : \langle N, X \rangle = c\} \end{aligned}$$

and $\bar{\varepsilon} := \min\{\varepsilon \geq 0 : B_{r*}^{\varepsilon} \cap H \cap \mathcal{C} \neq \emptyset\}$. Theorem 2 states, if $\sigma_r(N + D^*) \neq \sigma_{r+1}(N + D^*)$ then $B_{r*}^{\bar{\varepsilon}} \cap H \cap \mathcal{C}$ consists of a single element. This can also be understood geometrically with the help of the following Lemma, which generalizes the corresponding result for the nuclear norm and $r = 1$.

LEMMA 2. *The set of the extreme points of the unit-ball B_{r*}^1 is*

$$E := \{X \in \mathbb{R}^{n \times m} : \|X\|_F = 1, \text{rank}(X) \leq r\}.$$

Hence, $B_{r*}^1 = \text{conv}(E)$, where $\text{conv}(\cdot)$ denotes the convex hull.

Proof. By the triangle inequality and (3) in Lemma 1 it follows that $\text{conv}(E) \subset B_{r*}^1$ (see Figure 2a). Moreover, by (1) in Lemma 1 it holds that

$$(9) \quad \forall N \in \mathbb{R}^{n \times m} : \sup_{M \in \text{conv}(E)} \langle N, M \rangle = \|N\|_r = \sup_{M \in B_{r*}^1} \langle N, M \rangle.$$

Since $\text{conv}(E)$ and B_{r*}^1 are closed sets, (9) holds if and only if $B_{r*}^1 = \text{conv}(E)$. If a point $\bar{M} \in E$ is not an extreme point of E , then $\bar{M} = \sum_i \alpha_i M_i$, $\sum_i \alpha_i = 1$, $\alpha_i > 0$, $M_i \in K \setminus \{\bar{M}\}$ for all i . Thus, by the Cauchy-Schwarz inequality, we conclude that $\langle \bar{M}, M_i \rangle = 1$ for all i and $\bar{M} = M_i$, which is a contradiction. \square

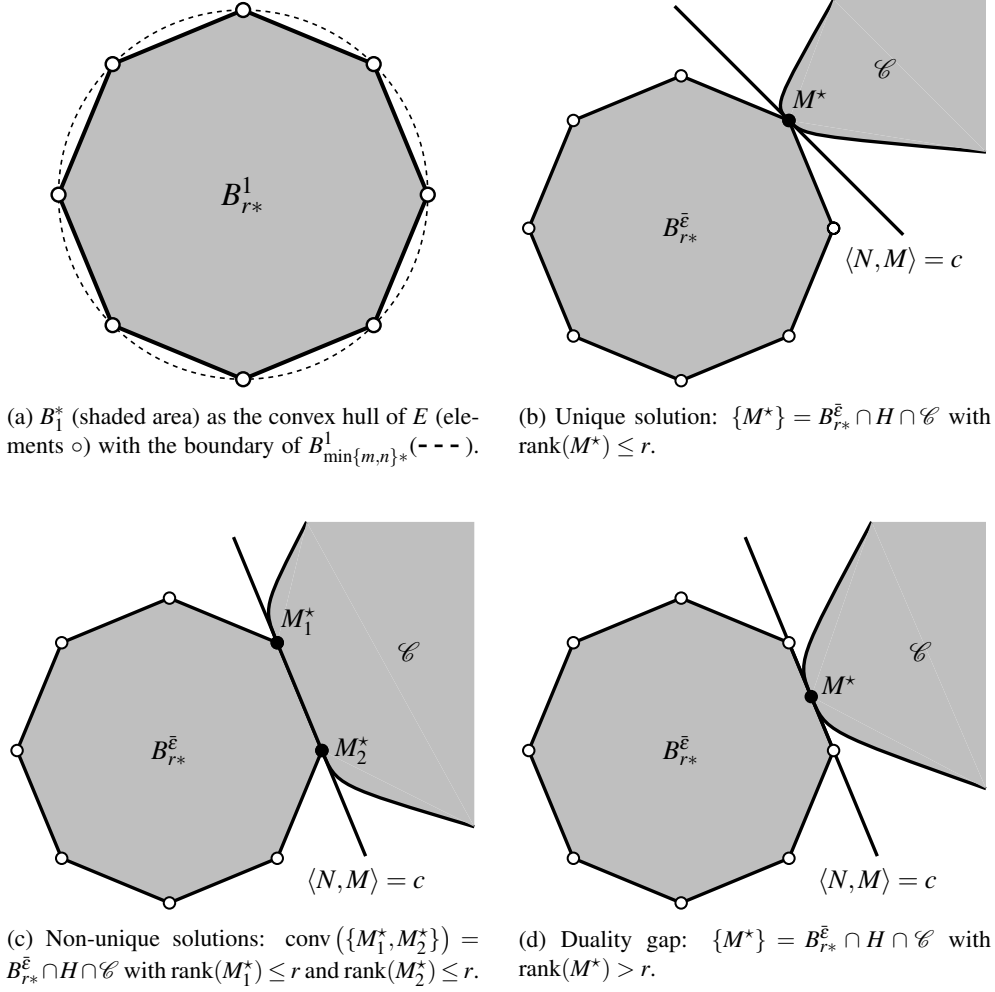


Figure 2: Schematic plots to visualize (8) geometrically.

Therefore, a geometric interpretation of $\sigma_r(N + D^*) \neq \sigma_{r+1}(N + D^*)$ is that the only intersection point of H and $B_{r*}^{\bar{e}} \cap \mathcal{C}$ is an extreme point of $B_{r*}^{\bar{e}}$ and \mathcal{C} (see Figure 2b). Hence, the case of $\sigma_r(N + D^*) = \sigma_{r+1}(N + D^*) \neq 0$ can occur if and only if H intersects $B_{r*}^{\bar{e}} \cap \mathcal{C}$ at several points (see Figure 2c and Subsection 6.1.2) or if there is a duality gap in (C) (see Figure 2d and Subsection 8.1). Finally notice that one can also use Lemma 2 as a definition of $\|\cdot\|_{r*}$. This has been done for vectors in [4] in an intention to generalize the ℓ_1 norm.

3.2. Real-valued r . In the following we will see that allowing r to be real-valued can be considered a regularization parameter. Unlike typical regularization methods (see Subsections 5.1 and 5.2), this parameter has a close relationship to the rank of the corresponding solutions. It suffices to discuss the case where Theorem 2 does not apply.

Let $r \in \mathbb{N}$ be such that $\sigma_r(N + D_r^*) = \sigma_{r+1}(N + D_r^*)$ and $\text{rank}(M_r^*) > r$, where we define

$$M_r^* := \left[\underset{M \in \mathbb{R}^{n \times m}}{\operatorname{argmin}} \frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle + g(M) \right], \quad D_r^* := \underset{D \in \mathbb{R}^{n \times m}}{\operatorname{argmin}} \left[g^*(-D) + \frac{1}{2} \|N + D\|_r^2 \right]$$

for $1 \leq r \leq \min\{m, n\}$. Assume that $\frac{1}{2} \|N - M_r^*\|_F^2 + g(M_r^*) > \frac{1}{2} \|N - M_{r+1}^*\|_F^2 + g(M_{r+1}^*)$ and $\text{rank}(M_{r+1}^*) > \text{rank}(M_r^*)$. Then one may be in the situation that M_r^* is an approximation of small rank but poor cost $\|N - M_r^*\|_F^2 + g(M_r^*)$. On the other hand, $\|N - M_{r+1}^*\|_F + g(M_{r+1}^*)$ may be acceptable while $\text{rank}(M_{r+1}^*)$ is too large. Thus a trade-off between M_r^* and M_{r+1}^* is desired. This can be achieved by letting r become a non-integer valued in the r norm. The r norm is then defined as

$$(10) \quad \|\cdot\|_r := \sqrt{\sum_{i=1}^{\lfloor r \rfloor} \sigma_i^2(\cdot) + (r - \lfloor r \rfloor) \sigma_{\lfloor r \rfloor}^2(\cdot)},$$

where $\lfloor r \rfloor := \max\{z \in \mathbb{Z} : z \leq r\}$ and $\lceil r \rceil := \min\{z \in \mathbb{Z} : z \geq r\}$. Observe that for $r \in \mathbb{N}$ and $\alpha \in [0, 1]$ we have

$$(11) \quad \|\cdot\|_{r+\alpha}^2 = (1 - \alpha) \|\cdot\|_r^2 + \alpha \|\cdot\|_{r+1}^2,$$

which means that $\|\cdot\|_{r+\alpha}^2$ is a convex combination of $\|\cdot\|_r^2$ and $\|\cdot\|_{r+1}^2$, thus indicating its usefulness in supplying the desired trade-off solution. Similar to Theorem 3, also with $r \in \mathbb{R}_{\geq 1}$ it remains true that $\text{rank}(M_r^*) \leq \lceil r \rceil + s$ if

$$(12) \quad \sigma_{\lceil r \rceil}(N + D_r^*) = \cdots = \sigma_{\lceil r \rceil + s}(N + D_r^*) > \sigma_{\lceil r \rceil + s + 1}(N + D_r^*).$$

Hence, allowing r to assume values in $\mathbb{R}_{\geq 1}$ may allow us to find solutions of both, lower rank and lower cost. Next let us have a closer look at the dependency of s on r in (12).

Let us define

$$F(r) := g^*(-D) + \frac{1}{2} \|N + D\|_r^2 + \frac{1}{2} \|N\|_F^2.$$

Using (2) in Lemma 1, we conclude that F is monotonically decreasing. In conjunction with the piecewise linearity in (11), it follows that F is convex and thus continuous. From Berge's Maximum Theorem (see [6, p. 116] or [60, Theorem 9.17] for the convex case) it is known that the parameter depending set

$$\mathcal{C}^*(r) := \underset{D \in \mathbb{R}^{n \times m}}{\operatorname{argmin}} \left[g^*(-D) + \frac{1}{2} \|N + D\|_r^2 + \frac{1}{2} \|N\|_F^2 \right]$$

is upper hemicontinuous in r . This means that for all $r \in [1, \min\{m, n\}]$ and all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $t \geq 1$

$$(13) \quad |t - r| < \delta \Rightarrow \mathcal{C}^*(t) \subset \mathcal{B}_\varepsilon(\mathcal{C}^*(r)),$$

where $\mathcal{B}_\varepsilon(\mathcal{C}^*(r)) := \{X : \exists D \in \mathcal{C}^*(r) \text{ such that } \|X - D\|_F < \varepsilon\}$. For simplicity we assume that D_r^* is unique. By (13) and the continuity of the singular values (see [59, Corollary 4.9]), it follows that a sufficiently small increase of r does not increase s in (12). Therefore, as for the nuclear norm regularization, one often observes $\text{rank}(M_t^*)$, as a function of $t \in [r, r + 1]$, to look like a staircase (see Figure 9b in Section 8). Notice that a similar consideration can be done with

$$F(r) := \frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2 + g(M)$$

and

$$\mathcal{C}^*(r) := \operatorname{argmin}_{M \in \mathbb{R}^{n \times m}} \left[\frac{1}{2} \|M\|_{r^*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2 + g(M) \right].$$

In summary, real-valued r can be considered as a regularization parameter, similar to the regularization methods in Section 5.

4. Computability. This section is devoted to the computability aspects of the r^* approach. We show that the problems (A) and (B) can be formulated as SDPs if g is SDP-representable. Moreover, we compute the proximal-operators of f^* and f^{**} in Theorem 1. This allows us to solve (A) and (B) using a first order method such as Douglas-Rachford splitting. We further apply Douglas-Rachford to the original non-convex Problem 1. We show that if Theorem 2 applies, then its iterates coincide locally with the convex Douglas-Rachford.

4.1. SDP-representations. We start with an SDP-representation of the optimization problem

$$(14) \quad \min_{D \in \mathbb{R}^{n \times m}} \|N + D\|_r^2.$$

where $\|\cdot\|_r$ is defined as in (10) and $r \in [1, \min\{m, n\}]$. Let $T \in \mathbb{R}^{n \times n}$ be such that $T \succeq (N + D)(N + D)^T$, then $\operatorname{trace}(T) = \sum_{i=1}^n \sigma_i(T)$ and $\sigma_i(T) \geq \sigma_i^2(N + D)$ for all i such that $1 \leq i \leq \min\{m, n\}$ (see [39, Corollary 7.7.4.]). Therefore,

$$\|N + D\|_r^2 \leq \operatorname{trace}(T) - (\lceil r \rceil - r) \sigma_{\lceil r \rceil}(T) - \sum_{i=\lceil r \rceil+1}^n \sigma_i(T) \leq \operatorname{trace}(T) - (n - r) \sigma_n(T).$$

which is equivalent to

$$(15) \quad \|N + D\|_r^2 \leq \min_{T \succeq (N + D)(N + D)^T} \operatorname{trace}(T) - (n - r) \sigma_n(T).$$

In particular, equality in (15) can be achieved with

$$T^* := \sum_{i=1}^{\lceil r \rceil} \sigma_i^2(N + D) u_i u_i^T + \sigma_{\lceil r \rceil}^2(N + D) \sum_{i=\lceil r \rceil+1}^n u_i u_i^T$$

where $N + D = \sum_{i=1}^n \sigma_i(N + D) u_i v_i^T$ is an SVD of $N + D$. Furthermore, using the Schur-complement condition for positive semi-definiteness of $T - (N + D)(N + D)^T \succeq 0$ (see [39, Theorem 7.7.7.]) gives that (14) is equivalent to

$$\begin{aligned} & \text{minimize} && \operatorname{trace}(T) - \gamma(n - r) \\ & \text{subject to} && \begin{pmatrix} T & N + D \\ (N + D)^T & I \end{pmatrix} \succeq 0, \quad T \succeq \gamma I, \quad D \in \mathbb{R}^{n \times m}. \end{aligned}$$

Thus, if g is SDP-representable, then the dual of this optimization yields an SDP-formulation of (B) (see [9, 56] for $r = 1$). We get

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \operatorname{trace}(W) - \operatorname{trace}(N^T M) + g(M) \\ & \text{subject to} && \begin{pmatrix} I - P & M \\ M^T & W \end{pmatrix} \succeq 0, \quad P \succeq 0, \\ & && \operatorname{trace}(P) = m - r. \end{aligned}$$

Assuming that for $r \in \mathbb{N}$ it holds that $\sigma_r(N + D^*) \neq \sigma_{r+1}(N + D^*)$, the unique solution M^* to Problem 1 can be found directly, without computing the solution to (A).

4.2. Convex Douglas-Rachford. Many SDP-solvers are based on interior point methods (see [54, 62]). These solvers have good convergence properties, but the iteration complexity typically grows unfavorably with the problem dimension. In order to deal with problems of higher dimensions, it is often more desirable to look at first-order methods such as the Douglas-Rachford splitting algorithm (see [5, 19, 20, 46]). Let us recall the basic concept of this method. We want to determine a solution to

$$(16) \quad \underset{X}{\text{minimize}} \quad f(X) + g(X)$$

where $f, g : \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \cup \{\infty\}$ are closed and proper convex functionals. Then the Douglas-Rachford iteration is given by

$$(17a) \quad X^k = \text{prox}_{\gamma f}(Z^{k-1}),$$

$$(17b) \quad Y^k = \text{prox}_{\gamma g}(2X^k - Z^{k-1}),$$

$$(17c) \quad Z^k = Z^{k-1} + \rho(Y^k - X^k),$$

where $\gamma > 0$, $\rho \in (0, 2)$ and the proximal-operator is defined as

$$(18) \quad \text{prox}_{\gamma f}(Z) := \underset{X}{\text{argmin}} \left(f(X) + \frac{1}{2\gamma} \|X - Z\|_F^2 \right).$$

It is known that X^k and Y^k converge towards a minimizer of (16) (see [19, 20, 46]). A special case of these iterations is the well-known Alternating Direction Methods of Multipliers (ADMM) (see [8, 26, 29]). Note that the Douglas-Rachford splitting algorithm can also be applied to sums of more than two functionals f and g (see [17]).

Let g be as in (B) and assume that $\text{prox}_{\gamma g}(X)$ is easy to compute. In order to apply the Douglas-Rachford algorithm to (B) it remains to find $\text{prox}_{\gamma f}(Z)$ with

$$f(M) := \frac{1}{2} \|M\|_{r*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2.$$

We get

$$(19) \quad \begin{aligned} \text{prox}_{\gamma f}(Z) &= \underset{M \in \mathbb{R}^{n \times m}}{\text{argmin}} \left(\frac{1}{2} \|M\|_{r*}^2 - \langle N, M \rangle + \frac{1}{2} \|N\|_F^2 + \frac{1}{2\gamma} \|M - Z\|_F^2 \right) \\ &= \underset{M \in \mathbb{R}^{n \times m}}{\text{argmin}} \left(\frac{1}{2} \|M\|_{r*}^2 + \frac{1}{2\gamma} \|M - (\gamma N + Z)\|_F^2 + \langle Z, N \rangle \right) \\ &= \text{prox}_{\frac{\gamma}{2} \|\cdot\|_{r*}^2}(\gamma N + Z). \end{aligned}$$

Using the extended Moreau-decomposition (see [5, Theorem 14.3]) and Theorem 1, it holds that for all Z

$$\text{prox}_{\frac{\gamma}{2} \|\cdot\|_{r*}^2}(Z) + \gamma \text{prox}_{\frac{1}{2\gamma} \|\cdot\|_F^2}(\gamma^{-1} Z) = Z.$$

In combination with (19) we arrive at

$$(20) \quad \text{prox}_{\gamma f}(Z) = \gamma N + Z - \gamma \text{prox}_{\frac{1}{2\gamma} \|\cdot\|_F^2} \left(\frac{\gamma N + Z}{\gamma} \right).$$

Since $\text{prox}_{\frac{\gamma^{-1}}{2} \|\cdot\|_F^2} = \text{prox}_{\frac{1}{2\gamma} \|\cdot\|_F^2}$, it is sufficient to derive how to compute $\text{prox}_{\frac{\gamma}{2} \|\cdot\|_{r*}^2}$. This is done in Algorithm 1 on page 33 for $r \in [1, \min\{m, n\}]$. Explanatory derivations can be found in Subsection A.4. For integer-valued r similar derivations are presented in [22].

Finally, observe that if $r \in \mathbb{N}$ and

$$(21) \quad \sigma_r(\gamma N + Z) > (1 + \gamma^{-1}) \sigma_{r+1}(\gamma N + Z),$$

it follows from the derivations of $\text{prox}_{\frac{\gamma}{2}\|\cdot\|_F^2}$ (see Subsection A.4 and in particular (47)) that

$$\text{prox}_{\frac{1}{2\gamma}\|\cdot\|_F^2} \left(\frac{\gamma N + Z}{\gamma} \right) = \frac{\gamma N + Z}{\gamma} - \frac{1}{1 + \gamma} \text{svd}_r \left(\frac{\gamma N + Z}{\gamma} \right).$$

Therefore, (20) implies that

$$(22) \quad \text{prox}_{\gamma f}(Z) = \frac{1}{1 + \gamma} \text{svd}_r(\gamma N + Z).$$

We will use this fact in Subsection 4.4 to show a tight relationship with the non-convex Douglas-Rachford algorithm.

4.3. Douglas-Rachford limit point properties. A comparison between the Douglas-Rachford limit points and the optimality conditions for (A) and (B) (see Theorem 3), gives that all limit points $Z^* = \lim_{k \rightarrow \infty} Z^k$ of (17c) can be expressed as

$$(23) \quad Z^* = M^* + \gamma D^*,$$

where D^* and M^* are solutions to (A) and (B), respectively. Given M^* , Z^* and γ , this allows us to determine D^* . Moreover, by inspection of the Douglas-Rachford iterations, it can be shown that several known properties of the standard SVD-approximation remain true, if $\text{prox}_g(X)$ is preserving them.

PROPOSITION 4. *Let N and g be as in Problem 1. Then the following hold:*

- i. *Let $N \in \mathbb{S}$ and $\text{prox}_g(X) \in \mathbb{S}$ for all $X \in \mathbb{S}$. Then (A) and (B) have solutions $D^*, M^* \in \mathbb{S}$.*
- ii. *Let $Nv = 0$ and $\text{prox}_g(X)v = 0$ for all X with $Xv = 0$. Then (B) has a solution M^* such that $M^*v = 0$.*

In particular, the solution to Problem 1 preserves these properties, if (B) has a unique solution without a duality gap in (C).

Proof. Using [66, Theorem 2] it holds that $\text{prox}_{\frac{\gamma}{2}\|\cdot\|_{F^*}^2}(X)$ has the same singular vectors as X . Therefore, $\text{prox}_{\frac{\gamma}{2}\|\cdot\|_{F^*}^2}(X)$ preserves these properties and i. and ii. are proven by starting the Douglas-Rachford iterations for (B) with $Z^0 = 0$. The last claim follows with Proposition 3. \square

There are numerous reasonable choices of g such that Proposition 4 applies, a few examples will be discussed in Sections 6 to 8.

According to Theorem 2, $\sigma_r(N + D^*) \neq \sigma_{r+1}(N + D^*)$ is a sufficient condition for the uniqueness of a solution to (B). Note that without this assumption, a solution to Problem 1 does not necessarily preserve the properties in Proposition 4. This can be used to construct non-trivial examples where $\sigma_r(N + D^*) = \sigma_{r+1}(N + D^*)$ (see Subsection 6.1.2).

4.4. Non-convex Douglas-Rachford (NDR). Another approach to solve Problem 1 is to directly apply the Douglas-Rachford method to the non-convex problem

$$(24) \quad \min_{M \in \mathbb{R}^{n \times m}} \left[\frac{1}{2} \|N - M\|_F^2 + \chi_{\text{rank}(M) \leq r}(M) + g(M) \right].$$

This has the advantage that we are guaranteed to get a solution of desired rank, if the iterates converge. Recently, some local convergence guarantees for the non-convex Douglas-Rachford have appeared in the literature (see [34, 35, 55]). Here, we add to these findings by

showing that the non-convex Douglas-Rachford reduces locally to its convex counterpart if Theorem 2 applies. To this end, we start by deriving $\text{prox}_{\gamma\bar{f}}(Z)$ where

$$\bar{f}(M) := \frac{1}{2}\|N - M\|_F^2 + \chi_{\text{rank}(M) \leq r}(M).$$

We get

$$\begin{aligned} \text{prox}_{\gamma\bar{f}}(Z) &= \underset{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}}{\text{argmin}} \left(\frac{\gamma}{2}\|N - M\|_F^2 + \frac{1}{2}\|M - Z\|_F^2 \right) \\ &= \underset{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}}{\text{argmin}} \left(\frac{\gamma+1}{2}\|M\|_F^2 - \langle \gamma N + Z, M \rangle \right) \\ &= \underset{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}}{\text{argmin}} \left\| \frac{\gamma N + Z}{\gamma+1} - M \right\|_F^2. \end{aligned}$$

Hence, by Proposition 1

$$\frac{1}{1+\gamma} \text{svd}_r(\gamma N + Z) \in \text{prox}_{\gamma\bar{f}}(Z).$$

Next let D^* and M^* be solutions to (A) and (B), respectively. If the convex Douglas-Rachford iterations are applied to (B), then it follows by (23) that $Z^* = \gamma D^* + M^*$ is a limit point to (17c). Assuming that $\sigma_r(N + D^*) \neq \sigma_{r+1}(N + D^*)$, it holds that

$$\begin{aligned} (1 + \gamma^{-1})\sigma_{r+1}(\gamma N + Z^*) &= (1 + \gamma^{-1})\sigma_{r+1}(\gamma(N + D^*) + M^*) = (1 + \gamma)\sigma_{r+1}(N + D^*) \\ &< (1 + \gamma)\sigma_r(N + D^*) = \sigma_r(\gamma(N + D^*) + M^*) = \sigma_r(\gamma N + Z^*). \end{aligned}$$

By the continuity of the singular values (see [59, Corollary 4.9]), this allows us to conclude that (21) applies in a sufficiently small neighborhood of Z^* . Then (22) implies that for all Z within this neighborhood we get

$$\text{prox}_{\gamma\bar{f}}(Z) = \text{prox}_{\gamma f}(Z),$$

where $f(M) := \frac{1}{2}\|M\|_{F^*}^2 - \langle N, M \rangle + \frac{1}{2}\|N\|_{F^*}^2$. Hence, the convex and non-convex Douglas-Rachford iterations locally coincide. Furthermore, the Douglas-Rachford iterations cannot escape from this neighborhood, because the sequence $\|Z^* - Z^k\|_F$ of the convex Douglas-Rachford is known to be non-increasing (see [20]). Thus proving *local convergence* of the non-convex Douglas-Rachford when $\sigma_r(N + D^*) \neq \sigma_{r+1}(N + D^*)$.

Notice that if there is a zero duality-gap in (C), then it follows by Theorem 3 and (23) that the convex and non-convex Douglas-Rachford have limit points that correspond to a solution to Problem 1, even if $\sigma_r(N + D^*) = \sigma_{r+1}(N + D^*)$. We will see in Sections 6 to 8 that the non-convex Douglas-Rachford iterations can converge to these solutions. However, this may not be the case for all choices of Z^0 , since $\text{prox}_{\gamma\bar{f}}(Z)$ is not necessarily unique (see Subsection 6.1.2). Moreover, it is shown in Subsection 7.3 that if there is a duality-gap in (C), then the choice of γ can be crucial for the existence of a limit-point of the non-convex Douglas-Rachford.

Finally, observe that $\text{prox}_{\gamma\bar{f}}(Z)$ only requires to determine the dominant r singular values and singular vectors. Hence, sparse SVD solvers such as in [47] can be used to determine a dominant SVD and to gain more computational speed with large-scale problems. The same holds true for $\text{prox}_{\gamma f}(Z)$, where maybe a larger dominant SVD, but usually not a full SVD, needs to be determined.

5. Other approaches. In the following we compare the r^* approach to other methods that intend to solve Problem 1. These methods will also be used for numerical comparisons throughout the subsequent sections.

5.1. Nuclear Norm Regularization. One of the most widely used methods to approximate a solution to Problem 1 is the so-called nuclear norm regularization. It borrows techniques from sparse regularized regression or Lasso (see [61]), i.e. estimating a sparse solution \hat{x} to a linear system of equations $A\hat{x} \approx b$ by solving

$$\min_x \frac{1}{2} \|Ax - b\|_2^2 + \mu \|x\|_{\ell_1},$$

where $\|\cdot\|_2$ is the Euclidean norm, $\|x\|_{\ell_1} = \sum_{i \geq 1} |x_i|$ and $\mu \geq 0$ is a regularization parameter. A small number of non-zero singular values is equivalent to the matrix having low rank. Therefore, for given $N \in \mathbb{R}^{n \times m}$, a corresponding matrix version reads

$$(25) \quad \min_{M \in \mathbb{R}^{n \times m}} \frac{1}{2} \|N - M\|_F^2 + \mu \|M\|_{1*} + g(M),$$

where $g : \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \cup \{\infty\}$ is a given closed proper convex functional. The simplicity of this convexification as well as the results in [23, 24, 56] stimulated a big growth in applying this approach in many different areas (see [23, 24, 52, 56]). However, to get a specific rank, μ must be chosen a priori, which is often challenging. Commonly one assumes that the rank, as a function of μ , looks like a staircase, i.e. a large μ decreases the rank too much whereas a small μ may leave it too large. In order to find the best possible approximation, one usually likes to keep μ as small as possible, which on the other hand could end up in a costly search.

In general, even with the best possible choice of μ , this heuristic does not return an optimal solution to Problem 1. In particular, also for the simple case $g = 0$, one usually cannot choose μ such that the SVD-approximation, as required by Proposition 1, is obtained. Furthermore, there is no certificate for checking whether a solution is a minimizer of Problem 1.

5.2. Rank Regularization. Similar to the nuclear norm regularization, it has been suggested in [44, 43] to directly regularize on the rank, i.e.

$$\min_{M \in \mathbb{R}^{n \times m}} \left[\frac{1}{2} \|N - M\|_F^2 + \mu \text{rank}(M) + g(M) \right],$$

where $\mu \geq 0$ is a regularization parameter and $g : \mathbb{R}^{n \times m} \rightarrow \mathbb{R} \cup \{\infty\}$ a closed and proper convex functional. Since this problem is still non-convex, one needs to find a convex proxy of $f(M) := \frac{1}{2} \|N - M\|_F^2 + \mu \text{rank}(M)$. The conjugate and bi-conjugate functionals of f are given by (see [44, 43])

$$(26) \quad \begin{aligned} f^*(D) &= \frac{1}{2} \|N + D\|_F^2 - \frac{1}{2} \|N\|_F^2 - \frac{1}{2} \sum_{i=1}^{\min\{m,n\}} \min\{2\mu, \sigma_i^2(N + D)\}, \\ f^{**}(M) &= \frac{1}{2} \|M - N\|_F^2 + \frac{1}{2} \sum_{i=1}^{\min\{m,n\}} \left(2\mu - \max\{0, \sqrt{2\mu} - \sigma_i(M)\}^2 \right). \end{aligned}$$

Hence, by Fenchel-duality (see Lemma A.1 and Proposition A.3) it holds that

$$(27) \quad \min_{M \in \mathbb{R}^{n \times m}} [f(M) + g(M)] \geq - \min_{D \in \mathbb{R}^{n \times m}} [f^*(M) + g^*(-D)] = \min_{M \in \mathbb{R}^{n \times m}} [f^{**}(M) + g(M)].$$

Assume that there is no duality gap in (C) with solutions D^* and M^* to (A) and (B), respectively. Choosing $\frac{\sigma_F^2(N+D^*)}{2} \geq \mu \geq \frac{\sigma_{r+1}^2(N+D^*)}{2}$, it is readily seen that

$$f^*(D^*) = \frac{1}{2}\|N\|_F^2 - \frac{1}{2}\|N+D^*\|_r^2 + \mu r = \frac{1}{2}\|N-M^*\|_F^2 + \mu r + g(M^*),$$

where the last equality follows by Propositions 2 and 3. Hence,

$$\frac{1}{2}\|N-M^*\|_F^2 + \mu r + g(M^*) \geq - \min_{D \in \mathbb{R}^{n \times m}} [f^*(D) + g^*(-D)] \geq \frac{1}{2}\|N-M^*\|_F^2 + \mu r + g(M^*),$$

yielding equality in (27). Thus, this method obtains the same guaranteed optimal solutions as previously discussed for (A) and (B). Evidently, there is a strong relationship to Proposition 2 and Theorem 2. However, if there is a duality-gap, then the solutions may differ from those with non-integer valued $r \in [1, \min\{m, n\}]$ and it is unclear which method yields better results. Notice that, as in the case of real-valued r , Berge's maximum Theorem can be applied since f^* is continuous in μ . Again, this implies that the rank of the solutions is robust to small perturbations in μ . Nevertheless, there is no staircase behavior as one exhibits for real-valued r and the nuclear norm regularization (see Figure 9b in Subsection 8.1). Thus, since a good choice of the regularization parameter is usually a priori unknown, one is required to sweep over a large set of possible choices of μ .

The proximal-operator of f^{**} is computable (see [44, 43]) and therefore there are several first order optimization methods (see Subsection 4.2) that can be used to compute a minimizer of the right hand side in (27). Nevertheless, this usually limits one to choices of g that possess a cheaply computable proximal operator of g for even small dimensional examples. It is currently unknown if (26) is SDP-representable.

5.3. Projection-based methods. In the following let $g(M) = \chi_{\mathcal{C}}(M)$ be the characteristic function of a closed convex set \mathcal{C} . If the projection onto \mathcal{C} is computable, then there are several other heuristics of which a few are outlined next.

5.3.1. Lift-and-project Algorithm (LP). The idea of the so-called lift-and-project algorithm (see [16]) is to interchangeably perform a standard SVD-approximation of desired rank and project the result orthogonally onto the convex set \mathcal{C} , which again increases the rank. By starting with N as the first iterate, one hopes to keep the distance to N small. Naturally, this algorithm always returns the standard SVD-approximation of N if it lies within \mathcal{C} . Unfortunately, it is generally difficult to know whether the algorithm converges and if a possible limit point gives a satisfactory error (see [16]). However, if \mathcal{C} is a closed convex cone, then one can show that the Frobenius norm is decreased in every step and the convergence is guaranteed.

5.3.2. Alternating Least-Squares (ALS). All the so far considered approaches share the drawback that when implemented, their iterates usually need to converge in order to guarantee a feasible solution. The so-called alternating least-squares method is a way of overcoming this drawback by working with iterates that lie in \mathcal{C} and are of desired rank.

Given $V_0 \in \mathbb{R}^{r \times n} \setminus \{0\}$ such that $\{U \in \mathbb{R}^{m \times r} : UV_0 \in \mathcal{C}\} \setminus \{0\} \neq \emptyset$ one interchangeably solves

$$\begin{aligned} U_k &:= \operatorname{argmin}_{UV_{k-1} \in \mathcal{C}} \|N - UV_{k-1}\|_F^2, \\ V_k &:= \operatorname{argmin}_{U_k V \in \mathcal{C}} \|N - U_k V\|_F^2, \end{aligned}$$

with $k \geq 1$. Thus the rank constraint is explicitly taken into account by forming $U_k V_k$. Note that alternating least-squares without constraints converges for almost all V_0 to a standard SVD-approximation (see [58]). The results in Section 6 indicate that for certain choices of \mathcal{C} , this method also converges to an optimal solution if $\sigma_r(N + D^*) \neq \sigma_{r+1}(N + D^*)$. Moreover, there are examples where its solution attains the lower-bound of Proposition 2 even though $\sigma_r(N + D^*) = \sigma_{r+1}(N + D^*)$. In fact, this method often reproduces the same solutions as the non-convex Douglas-Rachford algorithm (see Subsection 4.4).

Nevertheless, in many cases ALS may not be a good choice since it is often unclear how to choose V_0 .

6. Non-negative low-rank approximation. A particularly well studied low-rank approximation problem is the case of preserving non-negativity constraints.

PROBLEM 2.

$$\begin{aligned} & \text{minimize} && \|N - M\|_F^2 \\ & \text{subject to} && M \in \mathbb{R}_{\geq 0}^{n \times m} \end{aligned}$$

where $\mathbb{R}_{\geq 0}^{n \times m} := \{X \in \mathbb{R}^{n \times m} : x_{ij} \geq 0\}$ and $N \in \mathbb{R}_{\geq 0}^{n \times m}$.

Note that this is the same as Problem 1 with $g = \chi_{\mathbb{R}_{\geq 0}^{n \times m}}$. The probably most well-known approach to this problem is the so-called non-negative matrix factorization (see [7, 41]). Given $N \in \mathbb{R}_{\geq 0}^{n \times m}$ one intends to find a solution to

$$\min_{\substack{U \in \mathbb{R}_{\geq 0}^{n \times r}, \\ V \in \mathbb{R}_{\geq 0}^{r \times m}}} \|N - UV\|_F^2.$$

Non-negative matrix factorization (NNMF) is often approximately solved by applying alternating least-squares (see [41] and Subsection 5.3.2). However, to require both, U and V to be non-negative might be very conservative, since Problem 2 only requires the product UV to be non-negative.

6.1. Examples. In the following we look at examples with a non-negativity constraint in order to illustrate several results that have been discussed in the previous sections.

6.1.1. Image compression. A common example in the literature (see [3, 21]) is to use the SVD for image compression. Given a grey-scale picture, one maps the pixels to a matrix of corresponding grey-scale values, typically integer values in $\{0, \dots, 255\}$, and performs a low-rank approximation of rank r . If r is sufficiently small, then the factors of the low-rank approximation are cheaper to store than the original matrix. Since the matrix is non-negative, it is very natural to keep this constraint intact. We apply all the methods that have been discussed so far to the Baboon-image in Figure 3a. A comparison among the relative errors of the methods as well as the normalized lower-bound obtained from (B), is shown in Figure 3b.

By the Perron-Frobenius Theorem (see [39, Theorem 8.4.4]) the rank-1 standard SVD-approximation is always non-negative. This reveals a major drawback of the nuclear norm heuristic for this problem, since it usually cannot recover standard SVD-approximations. Moreover, we observe that all the SVD-based methods produce results of similar quality. In fact, alternating least-squares (ALS), non-convex Douglas-Rachford (NDR) and the r^* approach give solutions that coincide numerically with the lower-bound, i.e. there is a zero duality gap for all ranks. The errors of the lift-and-project method are only slightly larger, but not visible in this plot.

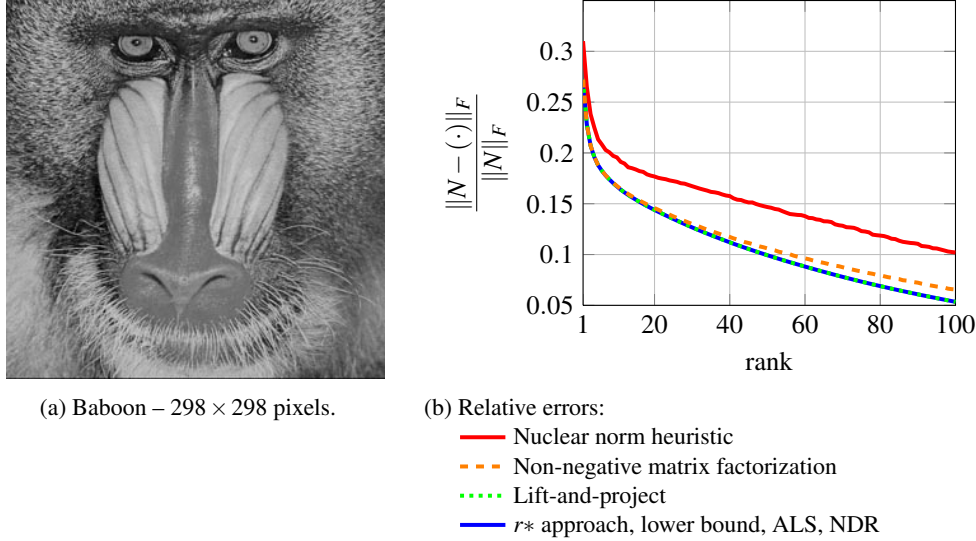


Figure 3: Non-negative Baboon-image approximation.

Non-negative matrix factorization (based on alternating least-squares), however, tends to produce larger errors with increasing rank. Overall, the nuclear norm heuristic performs significantly worse than any of the other methods.

6.1.2. Asymmetric optimal approximations. Let $N \in \mathbb{S} \cap \mathbb{R}_{\geq 0}^{n \times n}$ and D^* be a solution of (A) corresponding to Problem 2. According to Theorem 2 and Proposition 4 we know that if $\sigma_r(N + D^*) \neq \sigma_{r+1}(N + D^*)$, then $\text{svd}_r(N + D^*) \in \mathbb{S}$ is the unique solution to (B) and Problem 2. In the following we will see that preservation of symmetry may no longer be valid for an optimal non-negative approximation if $\sigma_r(N + D^*) = \sigma_{r+1}(N + D^*)$.

Consider Problem 2 with $r = 2$ and symmetric

$$N = \begin{pmatrix} \frac{\sqrt{5}-1}{2} & 1 & 3 \\ 1 & 4 & 1 \\ 3 & 1 & \frac{\sqrt{5}-1}{2} \end{pmatrix}.$$

A non-symmetric solution to this is

$$M^* = \begin{pmatrix} 0 & \frac{\sqrt{5}+1}{2} & \frac{\sqrt{5}+3}{2} \\ 2 & 3 & \frac{\sqrt{5}+1}{2} \\ 2 & 2 & 0 \end{pmatrix}.$$

Indeed, since N is symmetric, its singular values are given by the absolute value of its eigenvalues $\left\{ \pm \frac{7-\sqrt{5}}{2}, 3+\sqrt{5} \right\}$. Then with $\|N - M^*\|_F = \frac{7-\sqrt{5}}{2}$ and Proposition 1 we conclude that M^* and M^{*T} are optimal non-negative rank-2 approximations of N . Thus, by Corollary 1 it follows that $D^* = 0$ and $\sigma_2(N + D^*) = \sigma_3(N + D^*)$. Furthermore, it implies that M^* and M^{*T} are solutions to (B).

Since the solution set of a convex problem is convex, all points $\alpha M^* + (1 - \alpha) M^{*T}$ with $\alpha \in [0, 1]$ must be solutions to (B). However, $\text{rank}(\alpha M^* + (1 - \alpha) M^{*T}) = 3$ for all $\alpha \in (0, 1)$.

Thus we cannot expect to numerically find the rank-2 solutions by solving (B) (see Figure 2c). In particular, let either of the discussed Douglas-Rachford algorithms (see Subsections 4.2 and 4.4) be initialized with $Z^0 \in \mathbb{S}$. Then Proposition 4 implies that they may converge to a symmetric solution, which can be shown to be non-optimal for Problem 2.

Nevertheless, it is interesting to note that under random initialization, NDR and ALS often converge to an optimal solution.

7. Matrix Completion. Assuming that the entries of a matrix are only partially known, the so-called matrix completion problem asks when and how the unknown elements can be recovered. The low-rank assumption turned out to be suitable for theoretical developments as well as for many practical applications (see [11, 12, 13, 56, 67]). This leads to the following problem.

PROBLEM 3.

$$(28) \quad \begin{aligned} & \text{minimize} && \text{rank}(M) \\ & \text{subject to} && m_{ij} = z_{ij}, (i, j) \in \mathcal{J} \end{aligned}$$

where \mathcal{J} is an index set.

One of the most popular methods for solving Problem 3 is the technique introduced in [12]. It states that if $Z \in \mathbb{R}^{n \times n}$ then with high probability it is a solution to

$$(29) \quad \begin{aligned} & \text{minimize} && \|M\|_{1*} \\ & \text{subject to} && m_{ij} = z_{ij}, (i, j) \in \mathcal{J} \end{aligned}$$

if $\text{card}(\mathcal{J}) \geq Cn^{1.2}\text{rank}(Z)\log(n)$, where $\text{card}(\mathcal{J})$ denotes the cardinality of \mathcal{J} and C is a constant. Similar to that, it has been shown in [56] that (29) is able to detect a lowest rank solution. This means that one does not expect any other matrix of lower rank than Z having those partially known entries. Note that this formulation can be considered as a special case of Proposition 2 with $r = 1$, because

$$(30) \quad \min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\frac{1}{2} \|M\|_F^2 + g(M) \right] \geq - \min_{D \in \mathbb{R}^{n \times m}} \left[g^*(-D) + \frac{1}{2} \|D\|_r^2 \right] = \min_{M \in \mathbb{R}^{n \times m}} \left[\frac{1}{2} \|M\|_{r*}^2 + g(M) \right],$$

where $g(M) = \chi_{\mathcal{M}}(M)$ and $\mathcal{M} := \{M \in \mathbb{R}^{n \times m} : m_{ij} = z_{ij}, (i, j) \in \mathcal{J}\}$.

However, we suggest to keep the flexibility of r as a tuning parameter intact and consider instead

$$(31) \quad \begin{aligned} & \text{minimize} && \|M\|_{r*} \\ & \text{subject to} && m_{ij} = z_{ij}, (i, j) \in \mathcal{J}, \end{aligned}$$

where it is possible to sweep over real-valued $r \geq 1$. In Subsections 7.1, 7.2 and 7.4 we will see that this may significantly improve the quality of completion. Finally, let us see what is required of Z to be a solution to (31).

THEOREM 4. *Let $Z \in \mathbb{R}^{n \times m}$ with $r = \text{rank}(Z)$ and $\mathcal{J} \subset [1, \dots, n] \times [1, \dots, m]$. Then Z is a solution to (31) if and only if there exists $D^* \in \mathbb{R}^{n \times m}$ with $Z = \text{svd}_r(D^*)$ and $d_{ij}^* = 0$ for all $(i, j) \notin \mathcal{J}$.*

Proof. Let $g(M) = \chi_{\mathcal{M}}(M)$ and $\mathcal{M} := \{M \in \mathbb{R}^{n \times m} : m_{ij} = z_{ij}, (i, j) \in \mathcal{J}\}$. Then

$$g^*(D) = \sup_{M \in \mathcal{M}} \langle D, M \rangle < \infty \Leftrightarrow \forall (i, j) \in \mathcal{J} : d_{ij} = 0.$$

Hence, by Theorem 3 the existence of D^* such that $Z = \text{svd}_r(D^*)$ is necessary for Z to be a solution to (31).

Assume that there exists $D \in \mathbb{R}^{n \times m}$ such that $Z = \text{svd}_r(D)$ and $d_{ij} = 0$ for all $(i, j) \notin \mathcal{J}$. Then, by Theorem 3 it follows that $Z \in \partial_{\frac{1}{2}} \|D^*\|_F^2$. According to Proposition A.4 this is equivalent to $D^* \in \partial_{\frac{1}{2}} \|Z\|_{r*}^2$ and therefore for all $\tilde{Z} \in \mathcal{M}$ it holds that

$$\frac{1}{2} \|\tilde{Z}\|_{r*}^2 \geq \frac{1}{2} \|Z\|_{r*}^2 + \langle D^*, \tilde{Z} - Z \rangle = \frac{1}{2} \|Z\|_{r*}^2.$$

This shows the sufficiency and concludes the proof. \square

7.1. Some motivational examples. Next we want to demonstrate that $r > 1$ may help to complete matrices where $r = 1$ fails. To this end, consider the rank-2 matrices

$$Z_1 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad Z_2 = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad Z_3 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 4 \end{pmatrix}.$$

We would like to recover these matrices under the assumption that the zero entries are the only unknown ones. By Theorem 4 we know that (31) with $r = 2$ can do that. It can be shown, e.g. by Proposition 4, that solving (29) is equivalent to determining

$$(32) \quad \min_{t \in \mathbb{R}} \|Z_i(t)\|_{1*}, \quad i = 1, 2, 3$$

where

$$Z_1(t) := \begin{pmatrix} t & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad Z_2(t) = \begin{pmatrix} 2 & t & 1 \\ t & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad Z_3(t) = \begin{pmatrix} t & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 4 \end{pmatrix}.$$

First we show that finding the lowest rank solution may not be sufficient to recover the true matrix. In case of Z_1 we get that $\text{rank}(Z_1(t)) = 1$ if and only if $t = 1$. Moreover, for $u := (-1 \ 0.5 \ 0.5)^T$ it holds that $\|uu^T\|_F < \|Z_1(1)\|_F$ and $Z_1(1)u = 0$. Hence, as required by Theorem 4, $D^* = Z_1(1) - uu^T$ guarantees that $Z_1(1)$ is the unique solution to (29) and therefore the nuclear norm heuristic does not recover Z_1 .

Next we show that non-uniqueness in (29) is another issue that can be avoided with the proposed approach in Theorem 4. Since $Z_2(t)$ is symmetric, it holds that

$$\|Z_2(t)\|_{1*} \geq \text{trace}(Z_2(t)) \equiv 5$$

with equality if and only if $Z_2(t) \succeq 0$. It is readily seen that $Z_2(t) \succeq 0$ if and only if $t \in [0, 2]$, which implies that all of these points are solutions to the nuclear norm heuristic (29). However, a numerical solver for (29) does not necessarily determine Z_2 .

Finally, observe that the nuclear norm heuristic does not always determine the lowest rank solution. It holds that $\text{rank}(Z_3(t)) \geq 2$ with equality if and only if $t = 0$. Moreover, it can be verified that $\|Z_3\|_{1*} > \|Z_3(0.1)\|_{1*}$. Thus Z_3 is not a solution to (29).

These examples show that additional knowledge about the true rank as well as the minimality in the Frobenius norm (see (30)) can be utilized with $\|\cdot\|_{r*}$ to possibly gain better completion. The following subsections will demonstrate the same behavior for a larger example and a practical application.

Finally, note that in view of (25) one may also consider

$$(33) \quad \begin{aligned} & \text{minimize} && \frac{1}{2} \|M\|_F^2 + \mu \|M\|_{1*} \\ & \text{subject to} && m_{ij} = z_{ij}, \quad (i, j) \in \mathcal{J}, \end{aligned}$$

where one sweeps over $\mu \geq 0$. This is a strategy that has been discussed earlier in [10]. Applied to the previous examples, this approach is also able to recover Z_1 , Z_2 and Z_3 with $\mu = 0$. Nevertheless, we will see that generally there may not be any μ that leads to a low-rank solution.

7.2. Numerical Example. The following example is intended to show a numerical comparison among (31) and (33). Let $Z = \text{svd}_5(H)$ where $H \in \mathbb{R}^{10 \times 10}$ is a Hankel-matrix of the following structure

$$H = \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ & 1 & \ddots & & 0 \\ & & \ddots & & \vdots \\ 1 & & & & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Moreover, let the index-set of the known entries be $\mathcal{J} = \{(i, j) : z_{ij} > 0\}$.

Figure 4 shows the relative completion errors as well as the obtained ranks of the solutions M_r^* of (31) for different integer-valued r . The corresponding results for M_μ^* , obtained by sweeping over $\mu \geq 0$ in (33), are presented in Figure 5.

The solution to the nuclear norm heuristic M_1^* ($r = 1$), gives the worst completion error and full rank. Notice that $n^{1.2} \text{rank}(Z) \log(n) \gg \text{card}(\mathcal{J}) = 22$ which is why one cannot expect that the example lies within the scope of this method. In contrast, M_5^* ($r = 5$) recovers the true matrix and is a sweet spot among all solutions. In fact, this is also guaranteed by Theorem 4 since $\mathcal{J} \subset \{(i, j) : h_{ij} = 0\}$. Furthermore, there is no μ such that $\text{rank}(M_\mu^*) < 10$.

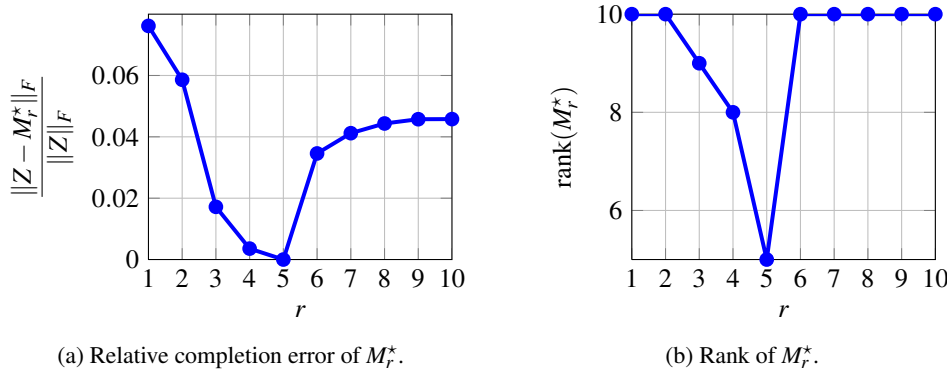
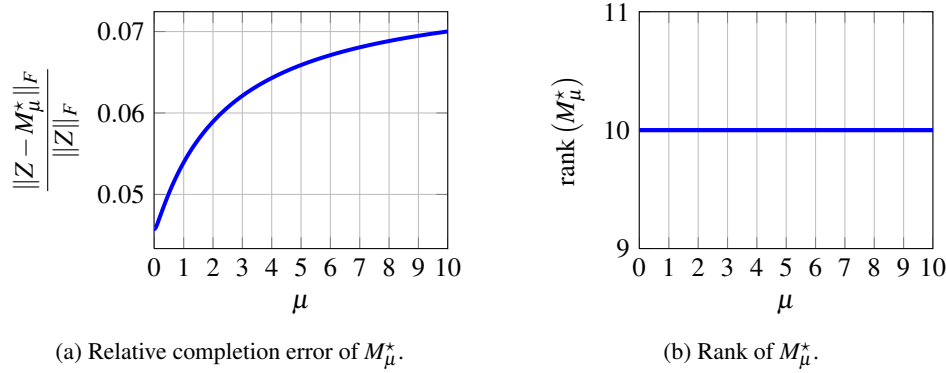


Figure 4: Relative completion error and ranks obtained with (31) for different values of r .

7.3. Example: Non-convex Douglas-Rachford. In the following we use Theorem 4 to construct examples where the nuclear norm heuristic as well as the r^* -approach fail to determine a solution to Problem 3. This helps to understand why the non-convex Douglas-Rachford (see Subsection 4.4) may still be able to find those solutions and that, unlike in the convex Douglas-Rachford, the choice of γ is crucial.

First note that the existence of D^* in Theorem 4 is equivalent to having an $R \in \mathbb{R}^{n \times m}$ such

Figure 5: Relative completion error and ranks obtained with (33) for different values of μ .

that

$$(34a) \quad R^T Z = 0, \quad Z R^T = 0, \quad \sigma_1(R) \leq \sigma_r(Z),$$

$$(34b) \quad z_{ij} + r_{ij} = 0 \text{ for all } (i, j) \notin \mathcal{J}.$$

Let us define for $t \in [-1, 1]$ the following unitary rank-1 matrix

$$Z := \begin{pmatrix} t \\ \sqrt{1-t^2} \end{pmatrix} \begin{pmatrix} t & \sqrt{1-t^2} \end{pmatrix} = \begin{pmatrix} t^2 & t\sqrt{1-t^2} \\ t\sqrt{1-t^2} & 1-t^2 \end{pmatrix}.$$

Correspondingly, all $R \in \mathbb{R}^{2 \times 2}$ that fulfill (34a) are given by

$$R = k \begin{pmatrix} \sqrt{1-t^2} \\ -t \end{pmatrix} \begin{pmatrix} \sqrt{1-t^2} & -t \end{pmatrix} = \begin{pmatrix} 1-t^2 & -t\sqrt{1-t^2} \\ -t\sqrt{1-t^2} & t^2 \end{pmatrix}$$

where $k \in [-1, 1]$. If $\mathcal{J} = \{(1, 2), (2, 1), (2, 2)\}$, it follows that (34b) can be satisfied if and only if $t^2 \leq \frac{1}{2}$. Hence, despite the fact that the solution to Problem 3 is unique, neither the nuclear norm heuristic nor the r^* approach is able to determine it if $t^2 > \frac{1}{2}$.

Next let us look at the limit-points of the non-convex Douglas-Rachford. Assume X^* , Y^* and Z^* are limit-points of the iterations (17a)–(17c) of the non-convex Douglas-Rachford applied to

$$\min_{\substack{M \in \mathbb{R}^{n \times m} \\ \text{rank}(M) \leq r}} \left[\frac{1}{2} \|M\|_F^2 + g(M) \right]$$

with $g(M) = \chi_{\mathcal{M}}(M)$ and $\mathcal{M} := \{M \in \mathbb{R}^{n \times n} : m_{ij} = z_{ij}, (i, j) \in \mathcal{J}\}$.

By (17a) and (17c) it follows that $X^* = \frac{1}{1+\gamma} \text{svd}_r(Z^*) = Y^*$. Moreover, if $R^* := Z^* - \text{svd}_r(Z^*)$ then (17b) implies that

$$\gamma x_{ij}^* + r_{ij}^* = 0 \text{ for all } (i, j) \notin \mathcal{J}.$$

Therefore, the non-convex Douglas-Rachford has a limit-point at $X^* \in \mathcal{M}$ if and only if there exists $R \in \mathbb{R}^{n \times m}$ such that

$$\begin{aligned} R^T X^* &= 0, \quad X^* R^T = 0, \quad \sigma_1(R) \leq (1 + \gamma^{-1}) \sigma_r(X^*), \\ x_{ij}^* + r_{ij} &= 0 \text{ for all } (i, j) \notin \mathcal{J}, \end{aligned}$$

where the inequality follows by the fact that $(1 + \gamma)\sigma_r(X^*) = \sigma_r(Z^*) \geq \sigma_1(R^*) = \gamma\sigma_1(\gamma^{-1}R^*)$. Thus, in the non-convex Douglas-Rachford R is allowed to be $(1 + \gamma^{-1})$ -times as large as in (34a). For sufficiently small $\gamma > 0$ it follows that all rank- r elements in \mathcal{M} are limit-points. Applied to Z and \mathcal{S} from above, we conclude that for all $t \in [-1, 1]$ there exists $\gamma > 0$ such that non-convex Douglas-Rachford has a limit-point at Z . Indeed, numerical computations suggest that the algorithm also converges to Z .

Nevertheless, this also demonstrates that a wrong choice of γ may prevent the existence of a limit-point.

7.4. Covariance completion. Consider

$$\dot{x}(t) = Ax(t) + Bu(t),$$

where $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times m}$ with $m \leq n$ and $u(t)$ is a zero-mean stationary stochastic process. For Hurwitz A and reachable (A, B) it has been shown (see [27, 28]) that the following are equivalent:

- i. $X := \lim_{t \rightarrow \infty} \mathbf{E}(x(t)x^T(t)) \succeq 0$ is the steady-state covariance matrix of $x(t)$, where $\mathbf{E}(\cdot)$ denotes the expected value.
- ii. $\exists H \in \mathbb{R}^{m \times n} : AX + XA^T = -(BH + H^T B^T)$.
- iii. $\text{rank} \begin{pmatrix} AX + XA^T & B \\ B^T & 0 \end{pmatrix} = \text{rank} \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$.

In particular, $H = \frac{1}{2}\mathbf{E}(u(t)u^T(t))B^T$ if u is white noise. In [15, 45, 67, 68, 69] the problem of unknown B and only partially known X has been addressed as follows.

PROBLEM 4.

$$\begin{aligned} & \text{minimize} && \text{rank}(M) \\ & \text{subject to} && \hat{x}_{ij} = x_{ij}, (i, j) \in \mathcal{S} \\ & && A\hat{X} + \hat{X}A^T = -M \\ & && \hat{X} \succeq 0. \end{aligned} \tag{35}$$

The problem has been solved in the same manner as Problem 3 i.e. by convexifying $\text{rank}(M)$ with $\|M\|_{1*}$. However, since some practical examples only supply up to $2n$ known entries of specific structure (see [67, 68, 69]), it is not surprising that the quality of completion is often not satisfactory.

Instead, in [32] its generalization as in (31) is considered, i.e.

$$\begin{aligned} & \text{minimize} && \|M\|_{r*} \\ & \text{subject to} && \hat{x}_{ij} = x_{ij}, (i, j) \in \mathcal{S} \\ & && A\hat{X} + \hat{X}A^T = -M \\ & && \hat{X} \succeq 0, \end{aligned} \tag{36}$$

where it is possible to sweep over $r \geq 1$. Again, one may also consider

$$\begin{aligned} & \text{minimize} && \frac{1}{2}\|M\|_F^2 + \mu\|M\|_{1*} \\ & \text{subject to} && \hat{x}_{ij} = x_{ij}, (i, j) \in \mathcal{S} \\ & && A\hat{X} + \hat{X}A^T = -M \\ & && \hat{X} \succeq 0, \end{aligned} \tag{37}$$

while sweeping over $\mu \geq 0$.

7.4.1. Example: Discretized Heat-Equation. Let us illustrate these approaches by a numerical comparison. Consider the two-dimensional heat-equation

$$\dot{T} = \Delta T = \frac{\partial^2}{\partial x^2} T + \frac{\partial^2}{\partial y^2} T$$

on the unit-square. Finite difference discretization on a uniform grid with step size $h = \frac{1}{N+1}$ gives

$$\Delta T_{ij} \approx -\frac{1}{h^2} (4T_{ij} - T_{i+1,j} - T_{i,j+1} - T_{i-1,j} - T_{i,j-1}),$$

where T_{ij} are the temperatures of the inner grid points as indicated in Figure 6. By letting the

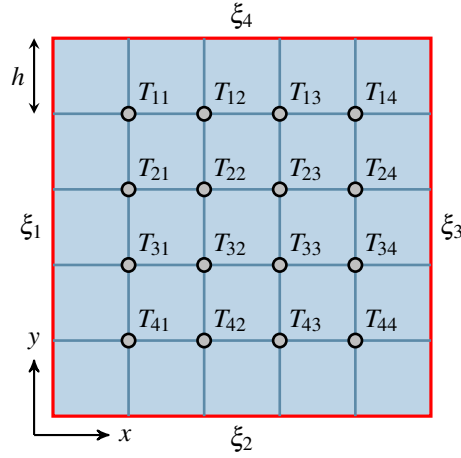


Figure 6: Discretized grid on the unit square with inputs ξ_1, \dots, ξ_4 .

boundaries of the unit-square be the inputs, we receive a linear system

$$(38) \quad \dot{x}(t) = \frac{1}{h^2} A x(t) + \frac{1}{h^2} B \xi(t)$$

where $A \in \mathbb{R}^{N^2 \times N^2}$ is the Poisson-matrix and $B = [b_{ij}] \in \mathbb{R}^{N^2 \times 4}$ with $b_{ij} = 0$ except for the following cases:

$$\begin{aligned} b_{i1} &:= 1, & \text{for } i = 1, 2, \dots, N \\ b_{i2} &:= 1, & \text{for } i = N, 2N, \dots, N^2 \\ b_{i3} &:= 1, & \text{for } i = N(N-1) + 1, N(N-1) + 2, \dots, N^2 \\ b_{i4} &:= 1, & \text{for } i = 1, N+1, \dots, N(N-1) + 1. \end{aligned}$$

Moreover, let $\xi(t)$ be generated by a low-pass filtered white-noise signal $w(t)$ with unit covariance $\mathbf{E}(w(t)w(t)^T) = I$ and

$$\dot{\xi}(t) = -\xi(t) + w(t).$$

As mentioned before, the extended covariance matrix

$$X_e := \mathbf{E}(x_e x_e^T) = \begin{pmatrix} X & X_{x\xi} \\ X_{\xi x} & X_\xi \end{pmatrix} \text{ with } x_e := \begin{pmatrix} x(t) \\ \xi(t) \end{pmatrix}$$

is then determined by

$$A_e X_e + X_e A_e^T = -B_e B_e^T,$$

where $A_e := \begin{pmatrix} A & B \\ 0 & -I \end{pmatrix}$, $B_e := \begin{pmatrix} 0 \\ I \end{pmatrix}$ and X is the steady-state covariance matrix of $x(t)$.

In the following we assume that only the first and third input channels are used, i.e. we remove the second and fourth columns from B and adjust A_e , B_e and $\xi(t)$, accordingly. An interpolated colormap of X is shown in Figure 7a, where the black lines indicate the known entries. Figure 7b displays the relative completion error of the solutions obtained by (36) and (37) with dependency on r and μ . We observe that the error obtained by (36) in $r = 2$ is the smallest and in fact it is of rank 2. This implies that there is no duality-gap. In contrast, the best solution that originates from (37) (with $\mu = 4.23$) is of rank 3 and has an error that is about 1.5 times as large. Figure 8 illustrates these differences through the interpolated colormaps.

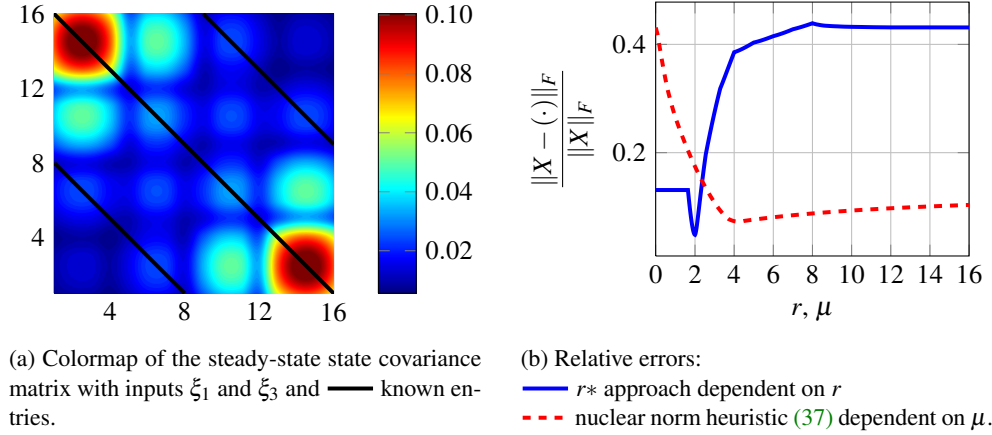


Figure 7: Interpolated colormap of the steady-state covariance matrix and relative errors dependent on r and μ obtained by (36) and (37).

8. Hankel matrices. In the field of system and control, the rank of a Hankel operator/matrix is crucial, since it determines the complexity (order) of a linear system. By that it tells how costly it is to simulate a system or to implement a controller (see [3, 70]). For this reason, much focus was put into areas such as model order reduction. Even though the celebrated Adamyan-Arov-Krein Proposition (see [3, 53]) answers the question of optimal low-rank approximation of infinite dimensional Hankel operators, the following finite dimensional case still remains open.

PROBLEM 5.

$$\begin{aligned} & \underset{M}{\text{minimize}} && \|N - M\|_F^2 \\ & \text{subject to} && \text{rank}(M) \leq r \\ & && M \in \mathcal{H}, \end{aligned}$$

where $N \in \mathcal{H} := \{H : H \text{ is Hankel}\}$.

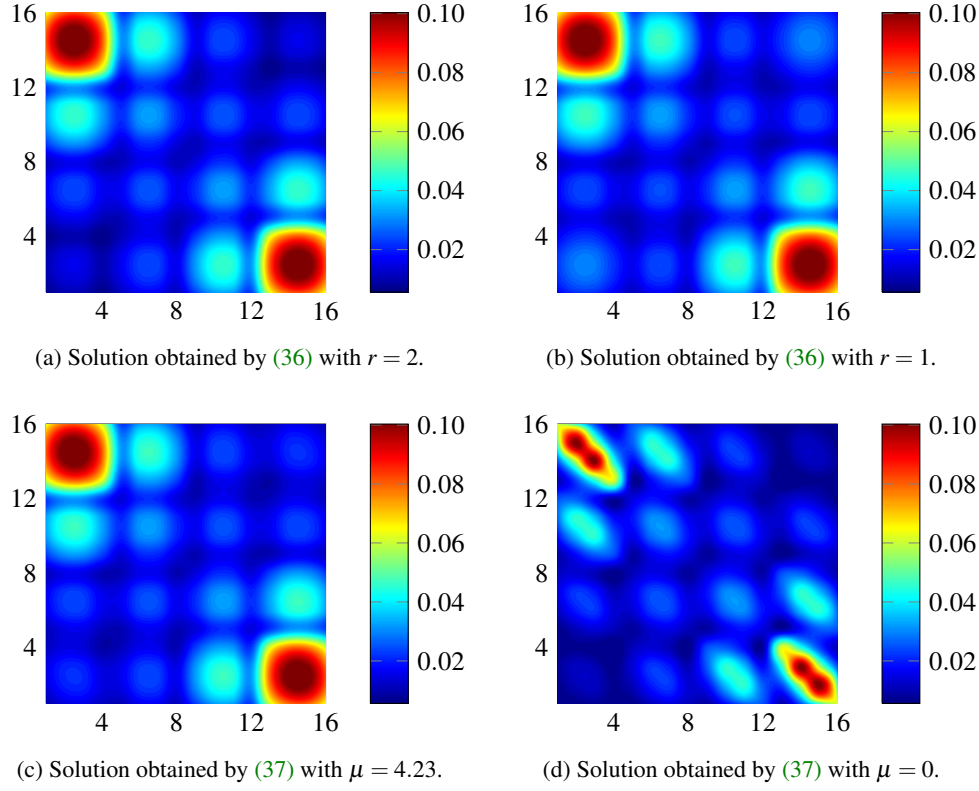


Figure 8: Interpolated colormaps of the completed covariance matrices obtained by (36) and (37).

Only the optimal rank-1 approximation in case of the spectral norm $\|\cdot\|_1$ has been determined in [2]. Moreover, for so-called linear externally positive systems the problem of non-negativity preserving Hankel-operator approximation has been considered in [30].

8.1. Numerical Example. In the following we compare the r^* approach with the regularization methods in Subsections 5.1 and 5.2 as well as the lift-and-project algorithm from Subsection 5.3.1. To this end, let $N \in \mathbb{R}^{10 \times 10}$ be the following Hankel matrix

$$N = \begin{pmatrix} 1 & 2 & \cdots & 9 & 10 \\ 2 & & \ddots & & 9 \\ \vdots & & & \ddots & \vdots \\ 9 & & & & 2 \\ 10 & 9 & \cdots & 2 & 1 \end{pmatrix}$$

The relative errors together with the relative lower bound are shown in Figure 9a. For $r = 1, \dots, 4$ there is a zero duality gap and therefore the lower bound is achieved by the rank-regularization method (see Subsection 5.2) as well as the non-convex Douglas-Rachford and the r^* norm. Moreover, even when Theorem 2 cannot guarantee a zero duality gap, it appears that those methods and the lift-and-project algorithm are close to the lower bound

and outperform the nuclear norm heuristic. Nonetheless, in order to get these (sub-optimal) solutions, we had to sweep over real-valued r and μ , respectively. The dependency of the rank on these parameters is displayed in Figure 9b.

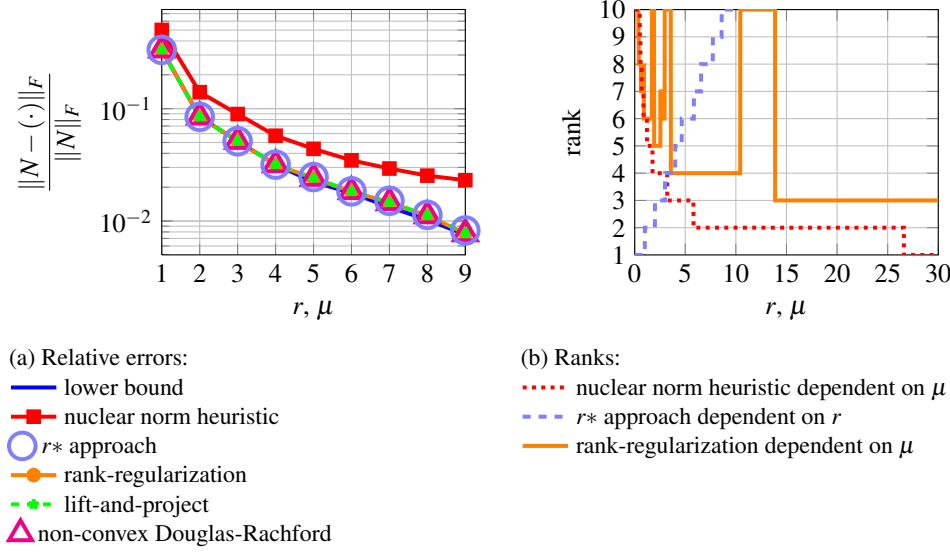


Figure 9: Relative error and rank dependency on r and μ – Hankel matrix preservation.

In contrast to the nuclear norm regularization and the r^* norm, that show the expected staircase behavior, the rank-regularization method seems to exhibit a non-intuitive oscillation, which complicates the search for an optimal μ , e.g. via a bisection algorithm.

9. Multivariate reduced-rank regression. In multivariate linear regression one wants to estimate a regression matrix $C \in \mathbb{R}^{n \times m}$ assuming the underlying linear model

$$Y = CX + E$$

where $Y \in \mathbb{R}^{n \times K}$ is a matrix with K measurements of n response variables, $X \in \mathbb{R}^{m \times K}$ are the corresponding predictor variables and $E \in \mathbb{R}^{n \times K}$ is Gaussian white-noise. Assuming that $\text{rank}(X) = m < T$ one can determine the well-known least-squares estimator

$$\hat{C} = YX^T(XX^T)^{-1},$$

which is a minimizer of $\min_C \|Y - CX\|_F^2$. Let \hat{c}_k and y_k denote the k -th row of \hat{C} and Y , respectively, then

$$\hat{c}_k = y_k X^T (XX^T)^{-1}$$

and therefore \hat{c}_k only depends on the k -th response variable y_k . Hence, the estimator does not account for possible correlations among the response variables.

In order to get estimators that include these correlations, one may restrict oneself to $\text{rank}(C) = r < \min\{m, n\}$ (see [40, 64]). Assuming that $C = AB$, where $A \in \mathbb{R}^{n \times r}$ and $B \in \mathbb{R}^{r \times m}$, a physical interpretation of this assumption on C can be given (see [64]). If X consists of information that is used to send T messages Y over r channels, then BX can be

considered a code for the information and ABX the decoded messages which are intended to be close to Y . Hence, given X , Y and r one would like to solve the problem

PROBLEM 6.

$$\begin{aligned} & \underset{C}{\text{minimize}} \quad \|Y - CX\|_F^2 \\ & \text{subject to} \quad \text{rank}(C) \leq r. \end{aligned}$$

Assuming that $\text{rank}(X) = m < K$, an explicit solution can be determined as follows. Let $X = U \begin{pmatrix} \Sigma & 0 \end{pmatrix} \begin{pmatrix} V_1 & V_2 \end{pmatrix}^T$ be an SVD of X with $\Sigma \in \mathbb{R}^{m \times m}$, then

$$\|Y - CX\|_F^2 = \|Y \begin{pmatrix} V_1 & V_2 \end{pmatrix} - \begin{pmatrix} CU\Sigma & 0 \end{pmatrix}\|_F^2 = \|YV_1 - CU\Sigma\|_F^2 + \|YV_2\|_F^2.$$

Hence, Problem 6 reduces to

$$\begin{aligned} (39) \quad & \underset{\tilde{C}}{\text{minimize}} \quad \|YV_1 - \tilde{C}\|_F^2 \\ & \text{subject to} \quad \text{rank}(\tilde{C}) \leq r. \end{aligned}$$

By Proposition 1 we know that a minimizer of (39) is given by $\text{svd}_r(YV_1)$ and therefore $\hat{C} = \text{svd}_r(YV_1)\Sigma^{-1}U^T$ is a solution to Problem 6. Observe that Problem 6 can also be stated as

$$\begin{aligned} & \underset{M}{\text{minimize}} \quad \frac{1}{2} \|Y - M\|_F^2 + \chi_{\mathcal{L}}(M) \\ & \text{subject to} \quad \text{rank}(M) \leq r, \end{aligned}$$

where $\mathcal{L} = \{M : M = CX \text{ for some } C \in \mathbb{R}^{n \times m}\}$ and thus fits into the scope of Proposition 2. Indeed, if $\text{rank}(X) = m$ then $\text{rank}(M) = \text{rank}(C)$ and solving

$$(40) \quad \underset{M}{\text{minimize}} \quad \frac{1}{2} \|M\|_{r^*}^2 - \langle Y, M \rangle + \chi_{\mathcal{L}}(M)$$

leads to the same solution as above if $\text{svd}_r(YV_1)$ is unique. This can be shown by considering the dual of (40). By Proposition 2 we get

$$\begin{aligned} & \underset{D}{\text{maximize}} \quad \frac{1}{2} \|Y + D\|_r^2 \\ & \text{subject to} \quad DX^T = 0, \end{aligned}$$

where $D^* = Y(V_1^T V_1 - I)$ is a feasible maximizer such that $\text{svd}_r(Y + D^*) = \text{svd}_r(YV_1 V_1^T) = \hat{C}X$ is a solution to (40).

Finally notice that further convex constraints on C can be added to (40) and by that more classes of regressors can be defined and computed.

10. Discussion and future developments. In this work, a method to determine optimal low-rank approximations with convex constraints has been studied. The main benefits of the r^* approach are that it is essentially parameter free, gives a certificate of optimality and does not depend on a particular initialization. Whereas factor based approaches such as alternating least-squares are also parameter free, they depended on its initialization and therefore are less applicable for general convex constraints, which can be handled with regularization approaches. The r^* approach combines the benefits of both approaches. Moreover, we have seen that it can be turned into a parameter dependent method, where unlike other approaches

the parameter has a clear relationship to the desired rank (see Subsection 3.2). As a result, a generalization of (29) to solve the matrix completion problem has been suggested. Furthermore, we have linked this approach to the rank-regularization method (see Subsection 5.2). Nevertheless, the r^* norm, in contrast to the rank-regularization method, is known to have an SDP-representation.

Since standard interior-point methods for SDPs are known to have iterations that grow unfavorably with dimension, the Douglas-Rachford splitting algorithm is used to gain computability for problems of larger dimensions. Based on that it was possible to show that several other useful properties known from the SVD-solution may be preserved (see Proposition 4). Moreover, it allowed us to show local convergence of the non-convex Douglas-Rachford if Theorem 2 applies. This motivates the overall usefulness of the non-convex Douglas-Rachford for solving Problem 1. This work is merely a starting point of investigating its power for solving the problems considered here. Further developments in this direction are likely to contribute to a better understanding of the duality-gap cases. One could start by linking the results in Subsection 4.4 to the known local convergence results in the vector case (see [35]). Moreover, it would be of great interest if, alike the nuclear norm minimization, r^* approach also fits into the scope of Matrix Manifold Optimization (see [1, 51, 63]).

The numerical examples in this paper indicate the superiority of the r^* approach and others over the nuclear norm heuristic. Since the r^* approach is as general as the nuclear norm heuristic, we suggest to use the r^* approach, instead. In fact, several other authors (see [4, 18, 22, 42, 50]) have recently used the r^* norm to replace the nuclear norm, i.e. another regularization with multiplicative parameter as in (33). However, neither taking advantage of its own regularization character nor the optimality.

Notice that despite the nice geometric interpretation (see Subsection 3.1), we were only able to guarantee a zero duality gap in simple cases such as Theorem 4. Investigating this further may lead to more deterministic guarantees.

Finally, observe that most of the results can be extended to Hilbert-Schmidt operators. As mentioned in Section 8, the finite dimensional case may be significantly more difficult than the infinite-dimensional one.

A. Appendix.

A.1. Unitarily invariant norms. The following results can be found e.g. in [39].

PROPOSITION A.1. *Let $A, B \in \mathbb{R}^{n \times m}$, then*

$$\langle A, B \rangle \leq \sum_{i=1}^{\min\{m,n\}} \sigma_i(A) \sigma_i(B).$$

COROLLARY A.1. *Let $A, B \in \mathbb{R}^{n \times m}$ then*

$$\sum_{i=1}^{\min\{m,n\}} \sigma_i(A) \sigma_i(B) = \max\{\langle A, UBV \rangle : U \text{ and } V \text{ are unitary}\}.$$

In the following we say that $g(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$ is a symmetric gauge function if and only if

- i. $g(\cdot)$ is a norm.
- ii. $\forall x \in \mathbb{R}^n : g(|x|) = g(x)$, where $|x|$ denotes the element-wise absolute value.
- iii. $g(Px) = g(x)$ for all permutation matrices P and all x .

PROPOSITION A.2. *$\|\cdot\|$ is a unitarily invariant norm on $\mathbb{R}^{n \times m}$ if and only if*

$$\|X\| = g(\sigma_1(X), \dots, \sigma_{\min\{m,n\}}(X)),$$

where g is a symmetric gauge function.

A.2. Convex Optimization. The following definitions and results from convex optimization (see [5, 37, 48, 57]) are used throughout the paper. In the following we assume that all functionals are defined on a real Hilbert space X with inner product $\langle \cdot, \cdot \rangle$. The *domain* of a functional f on X is defined as $\text{dom} f := \{x \in X : f(x) < \infty\}$.

DEFINITION A.1. Let $f : X \rightarrow \mathbb{R} \cup \{\infty\}$ be a functional with $\text{dom} f \neq \emptyset$, minorized by an affine functional i.e. $\exists (x^*, b) \in X \times \mathbb{R} : f(x) \geq \langle x, x^* \rangle - b$ for all $x \in X$. Then,

$$f^*(x^*) := \sup_{x \in X} [\langle x, x^* \rangle - f(x)]$$

is called its conjugate (dual) functional. Further, the bi-conjugate functional of f is defined as $f^{**} := (f^*)^*$.

DEFINITION A.2. A convex functional $f : X \rightarrow \mathbb{R} \cup \{\infty\}$ with $\text{dom} f \neq \emptyset$ is

- proper if $\text{dom} f \neq \emptyset$.
- closed if the epigraph $\{x : f(x) \leq t, x \in \text{dom} f\}$ is a closed set for all $t \in \mathbb{R}$.

It is known that $f^{**} = f$ if only if f is a closed and proper convex functional.

LEMMA A.1. Let $f, g : X \rightarrow \mathbb{R} \cup \{\infty\}$ be functionals as in Definition A.1. Then

$$(41) \quad \min_{x \in X} [f(x) + g(x)] \geq -\min_{x^* \in X} [f^*(x^*) + g^*(-x^*)].$$

PROPOSITION A.3. Let $f, g : X \rightarrow \mathbb{R} \cup \{\infty\}$ be closed and proper convex functionals. Assume that $\text{ri}(\text{dom} f) \cap \text{ri}(\text{dom} g) \neq \emptyset$ and $\text{ri}(\text{dom} f^*) \cap \text{ri}(\text{dom} g^*) \neq \emptyset$, where $\text{ri}(\cdot)$ denotes the relative interior. Then,

$$\min_{x \in X} [f(x) + g(x)] = -\min_{x^* \in X} [f^*(x^*) + g^*(-x^*)].$$

Moreover, if the minimum on the left is attained at some x_0 and the minimum on the right by some x_0^* , then

$$\begin{aligned} f^*(x_0^*) &= \langle x_0, x_0^* \rangle - f(x_0), \\ g^*(-x_0^*) &= \langle x_0, -x_0^* \rangle - g(x_0). \end{aligned}$$

DEFINITION A.3. Let $f : X \rightarrow \mathbb{R} \cup \{\infty\}$ be a functional. Then

$$\partial f(x_0) := \{x_0^* \in X : f(x) \geq f(x_0) + \langle x - x_0, x_0^* \rangle\}$$

is called the subdifferential of f at x_0 . Moreover, each $x_0^* \in \partial f(x_0)$ is referred to as a subgradient of f at x_0 .

PROPOSITION A.4. Let $f : X \rightarrow \mathbb{R} \cup \{\infty\}$ be a closed and proper convex functional. Then the following statements are equivalent:

- $x_0^* \in \partial f(x_0)$.
- $f^*(x_0^*) = \langle x_0, x_0^* \rangle - f(x_0)$.
- $x_0 \in \partial f^*(x_0^*)$.
- $f(x_0) = \langle x_0, x_0^* \rangle - f^*(x_0^*)$.

For $x \in \mathbb{R}^n$ and $r \in [1, n]$ we define $\|x\|_r := \sqrt[r]{g_r(x)}$ with

$$g_r(x) := \max\{x_{i_1}^2 + \dots + x_{i_{\lfloor r \rfloor}}^2 + (r - \lfloor r \rfloor)x_{i_{\lfloor r \rfloor}} : 1 \leq i_1 < i_2 < \dots < i_{\lfloor r \rfloor} \leq n\}.$$

The following Lemma on the subgradients of $\|\cdot\|_r$ has been shown in [18] for $r \in \mathbb{N}$. We simply extend it to the real-valued case.

LEMMA A.2. Let $r \in [1, n]$, $\bar{r} := \lceil r \rceil$ and $\sigma \in \mathbb{R}_{\geq 0}^n$ with

$$(42) \quad \sigma_1 \geq \dots > \sigma_{\bar{r}-t+1} = \dots = \sigma_{\bar{r}} = \dots = \sigma_{\bar{r}+s} > \dots \geq \sigma_n,$$

where $t = \bar{r}$ and $s = n - \bar{r}$ if $\sigma_1 = \sigma_{\bar{r}}$ and $\sigma_n = \sigma_{\bar{r}}$, respectively. Then $v \in \partial \|\sigma\|_r$ if and only if

- i. $1 \leq i \leq \bar{r} - t$: $v_i = \frac{\sigma_i}{\|\sigma\|_r}$.
- ii. $\bar{r} - t + 1 \leq i \leq \bar{r} + s$: $v_i = \tau_i \frac{\sigma_{\bar{r}}}{\|\sigma\|_r}$ with $0 \leq \tau_i \leq 1$, $\sum_{i=\bar{r}-t+1}^{\bar{r}+s} \tau_i = t - \bar{r} + r$.
- iii. $\bar{r} + s + 1 \leq i \leq n$: $v_i = 0$.

Proof. Let $r \in [1, n]$ and $\sigma \in \mathbb{R}_{\geq 0}^n$ as in (42). Then

$$\|\sigma\|_r = \max_{\substack{\mathcal{J} \subset \{1, \dots, n\} \\ \text{card}(\mathcal{J}) = \bar{r}}} g_{\mathcal{J}}(\sigma),$$

where $g_{\mathcal{J}}(\sigma) := \sqrt{\sum_{i \in \mathcal{J} \setminus \max(\mathcal{J})} \sigma_i^2 + (r - \lfloor r \rfloor) \sigma_{\max(\mathcal{J})}^2}$ and $\text{card}(\mathcal{J})$ denotes the cardinality of \mathcal{J} . Since $\|\sigma\|_r \neq 0$ it follows (see [38, Proposition 4.3.1]) that the sub-differentials of $\|\cdot\|_r$ evaluated at σ are given by

$$(43) \quad \partial \|\sigma\|_r = \text{conv} \{ \nabla g_{\mathcal{J}}(\sigma) : \mathcal{J} \subset \{1, \dots, n\}, \text{card}(\mathcal{J}) = \bar{r}, g_{\mathcal{J}}(\sigma) = \|\sigma\|_r \},$$

where ∇ denotes the gradient operator with respect to σ . Next we determine the gradient at these points where $\|\sigma\|_r = g_{\mathcal{J}}(\sigma)$. Then, by assumption (42) it holds that $\{1, \dots, \bar{r} - t\} \subset \mathcal{J}$ and therefore

- $1 \leq i \leq \bar{r} - t$: $\frac{\partial g_{\mathcal{J}}(\sigma)}{\partial \sigma_i} = \frac{\sigma_i}{\|\sigma\|_r}$.
- $i \in \mathcal{J} \cap \{\bar{r} - t + 1, \dots, \bar{r} + s\} \setminus \max(\mathcal{J})$: $\frac{\partial g_{\mathcal{J}}(\sigma)}{\partial \sigma_i} = \frac{\sigma_i}{\|\sigma\|_r}$.
- $i = \max(\mathcal{J})$: $\frac{\partial g_{\mathcal{J}}(\sigma)}{\partial \sigma_{\bar{r}}} = \frac{(r - \lfloor r \rfloor) \sigma_{\bar{r}}}{\|\sigma\|_r}$.
- $\bar{r} + s + 1 \leq i \leq n$: $\frac{\partial g_{\mathcal{J}}(\sigma)}{\partial \sigma_i} = 0$.

Thus, by (43) it holds that $v \in \partial \|\sigma\|_r$ if and only if

- i. $1 \leq i \leq \bar{r} - t$: $v_i = \frac{\sigma_i}{\|\sigma\|_r}$,
- ii. $\bar{r} - t + 1 \leq i \leq \bar{r} + s$: $v_i = \tau_i \frac{\sigma_{\bar{r}}}{\|\sigma\|_r}$ with $0 \leq \tau_i \leq 1$ and $\sum_{i=\bar{r}-t+1}^{\bar{r}+s} \tau_i = t - \bar{r} + r$,
- iii. $\bar{r} + s + 1 \leq i \leq n$: $v_i = 0$,

where the last part of the second condition follows from

$$\sum_{i \in \mathcal{J}} \frac{\partial g_{\mathcal{J}}(\sigma)}{\partial \sigma_i} = (t - \bar{r} + r) \frac{\sigma_{\bar{r}}}{\|\sigma\|_r}. \quad \square$$

From Lemma A.2 and [66, Theorem 2] the following Proposition follows in the same way as in [18] for $r \in \mathbb{N}$.

PROPOSITION A.5. Let $A \in \mathbb{R}^{n \times m}$, $r \in [1, \min\{m, n\}]$ and $\bar{r} := \lceil r \rceil$. Further, let an SVD of A be given by $A = \sum_{i=1}^{\min\{m, n\}} \sigma_i u_i v_i^T$ with $\sigma_{\bar{r}-t} \neq \sigma_{\bar{r}-t+1} = \dots = \sigma_{\bar{r}} = \dots = \sigma_{\bar{r}+s} \neq \sigma_{\bar{r}+s+1}$, where $t = \bar{r}$ and $s = \min\{m, n\} - \bar{r}$ if $\sigma_1 = \sigma_{\bar{r}}$ and $\sigma_{\min\{m, n\}} = \sigma_{\bar{r}}$, respectively. Then $M \in \partial \|A\|_r$ if and only if

$$M = \frac{1}{\|A\|_r} \left(\sum_{i=1}^{\bar{r}-t} \sigma_i u_i v_i^T + \sigma_{\bar{r}} (u_{\bar{r}-t+1} \quad \dots \quad u_{\bar{r}+s})^T (v_{\bar{r}-t+1} \quad \dots \quad v_{\bar{r}+s})^T \right),$$

where $T \succeq 0$, $\|T\|_{1*} = t + \bar{r} - r$ and $\|T\|_1 \leq 1$. In particular, if $\sigma_{\bar{r}} \neq \sigma_{\bar{r}+1}$ or $\sigma_{\bar{r}} = 0$ then $\text{rank}(M) \leq \bar{r}$.

A.3. Proof of Lemma 1.

Proof. Let $1 \leq r \leq q := \min\{m, n\}$ and

$$g(x_1, \dots, x_q) := \|\text{diag}(x_1, \dots, x_q)\|_r.$$

Then $\|\cdot\|_r$ is a unitarily invariant norm by Proposition A.2, because g is a symmetric gauge function. Now, let $M \in \mathbb{R}^{n \times m}$, then by Corollary A.1

$$\|M\|_r^2 = \max\{\langle M^T M, UPV \rangle : U \text{ and } V \text{ are unitary}\},$$

with $P := \begin{pmatrix} I_r & 0 \\ 0 & 0_{m-r} \end{pmatrix}$. If $M^T M = \sum_{i=1}^m \sigma_i(M) u_i u_i^T$ we can define a projection $P_r := \sum_{i=1}^r u_i u_i^T$ such that $\|M\|_r^2 = \langle P_r, M^T M \rangle$.

Since $\|\cdot\|_{r*}$ inherits the unitary invariance, we have

$$\begin{aligned} \|M\|_{r*} &= \|\Sigma\|_{r*} = \max_{\|X\|_r \leq 1} \langle \Sigma, X \rangle \leq \max_{\sum_i \sigma_i^2(X) = 1} \sum_{i=1}^q \sigma_i(M) \sigma_i(X) \\ &= \max_{\sum_i \sigma_i^2(X) \leq 1} \left[\sum_{i=1}^r \sigma_i(M) \sigma_i(X) + \sigma_r(X) \sum_{i=r+1}^q \sigma_i(M) \right], \end{aligned}$$

with $\Sigma := \text{diag}(\sigma_1(M), \dots, \sigma_q(M))$. The last inequality follows by Proposition A.1 and can be attained. Hence,

$$\|M\|_{*r} = \max_{\sum_i s_i^2 = 1} \sum_{i=1}^q \sigma_i(M) s_i \geq \max_{\sum_i s_i^2 = 1} \sum_{i=1}^r \sigma_i(M) s_i = \sum_{i=1}^r \sigma_i^2(M),$$

with equality if and only if $\text{rank}(M) \leq r$. \square

A.4. Derivation of $\text{prox}_{\frac{\gamma}{2}\|\cdot\|_r^2}(\cdot)$.

$$\text{prox}_{\frac{\gamma}{2}\|\cdot\|_r^2}(Z) = \underset{X}{\text{argmin}} \left(\frac{\gamma}{2} \|X\|_r^2 + \frac{1}{2} \|X - Z\|_F^2 \right).$$

which is equivalent to

$$\begin{aligned} X^* = \text{prox}_{\frac{\gamma}{2}\|\cdot\|_r^2}(Z) &\Leftrightarrow 0 \in \partial_X \left(\frac{\gamma}{2} \|X\|_r^2 + \frac{1}{2} \|X - Z\|_F^2 \right) \Big|_{X=X^*} \\ &\Leftrightarrow Z - X^* \in \gamma \|X^*\|_r \partial_X \|X\|_r \Big|_{X=X^*}. \end{aligned}$$

Let $\bar{r} := \lceil r \rceil$ and an SVD of X^* be given by $X^* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(X^*) u_i v_i^T$ such that

$$\sigma_{\bar{r}-t}(X^*) > \sigma_{\bar{r}-t+1}(X^*) = \dots = \sigma_{\bar{r}}(X^*) = \dots = \sigma_{\bar{r}+s}(X^*) > \sigma_{\bar{r}+s+1}(X^*),$$

where $t = \bar{r}$ and $s = n - \bar{r}$ if $\sigma_1(X^*) = \sigma_{\bar{r}}(X^*)$ and $\sigma_{\min\{m,n\}}(X^*) = \sigma_{\bar{r}}(X^*)$, respectively.

Further, let $U_2 := (u_{\bar{r}-t+1}, \dots, u_{\bar{r}+s})$ and $V_2 := (v_{\bar{r}-t+1}, \dots, v_{\bar{r}+s})$. Then, by Proposition A.5

$$Z = (1 + \gamma) \sum_{i=1}^{\bar{r}-t} \sigma_i(X^*) u_i v_i^T + \sigma_{\bar{r}}(X^*) U_2 (I + \gamma T) V_2^T + \sum_{i=\bar{r}+s+1}^{\min\{m,n\}} \sigma_i(X^*) u_i v_i^T$$

with $\|T\|_1 \leq 1$, $\|T\|_{1*} = t - \bar{r} + r$, $T \succeq 0$. Using [66, Theorem 2] it follows that Z has the same singular vectors as X^* and therefore $T = \text{diag}(T_{\bar{r}-t+1}, \dots, T_{\bar{r}+s})$ can be chosen to be diagonal. This gives

- i. $1 \leq i \leq \bar{r} - t$: $\sigma_i(X^*) = \frac{1}{1+\gamma} \sigma_i(Z)$.
- ii. $\bar{r} - t + 1 \leq i \leq \bar{r} + s$: $\sigma_i(X^*) = \frac{1}{1+\gamma T_i} \sigma_i(Z)$.
- iii. $\bar{r} + s + 1 \leq i \leq \min\{m, n\}$: $\sigma_i(X^*) = \sigma_i(Z)$.

Hence, the main task is to determine $s \geq 0$, $t \geq 1$ and $T \succeq 0$ such that

$$(44) \quad \sigma_{\bar{r}}(X^*) = \frac{\sigma_{\bar{r}-t+1}(Z)}{1 + \gamma T_{\bar{r}-t+1}} = \dots = \frac{\sigma_{\bar{r}+s}(Z)}{1 + \gamma T_{\bar{r}+s}},$$

where

$$(45) \quad \sum_{i=1}^s T_{\bar{r}-t+i} = t - \bar{r} + r \text{ and } T_{\bar{r}-t+i} \leq 1, \quad 1 \leq i \leq t + s$$

and

$$(46) \quad \sigma_{\bar{r}}(X^*) > \frac{\sigma_{\bar{r}-t+1}(Z)}{1 + \gamma} > \sigma_{\bar{r}+s+1}(Z).$$

Next we will show how s , t and T can be determined inductively. Clearly, there exists $T_{\bar{r}}, \dots, T_{\bar{r}+s_0}$ for some $s_0 \geq 0$, fulfilling (44) and (45) with $t = 1$ and $s = s_0$. However, if

$$\frac{1}{1 + \gamma} \sigma_{\bar{r}-1}(Z) \leq \frac{\sigma_{\bar{r}}(Z)}{1 + \gamma T_{\bar{r}}},$$

then requirement (46) is violated. Hence, $t = 0$ is not a feasible choice and we want to find the smallest possible t for which this requirement is met after constructing T . Let us assume that with $t = \tilde{t} - 1$ and $s = \tilde{s}_{\tilde{t}-1}$, there is no solution that satisfies all three conditions (44)–(46).

Then one can construct $T_{\bar{r}-\tilde{t}+1}, \dots, T_{\bar{r}+\tilde{s}_{\tilde{t}}}$ fulfilling (44) and (45) with $t = \tilde{t}$ and $s = \tilde{s}_{\tilde{t}}$, as follows:

Let $i \geq 2$ and $T_{\bar{r}-\tilde{t}+1}^{(i-1)}, \dots, T_{\bar{r}-\tilde{t}+i-1}^{(i-1)} \leq 1$ be determined such

$$\frac{\sigma_{\bar{r}-\tilde{t}+1}(Z)}{1 + \gamma T_{\bar{r}-\tilde{t}+1}^{(i-1)}} = \dots = \frac{\sigma_{\bar{r}-\tilde{t}+i-1}(Z)}{1 + \gamma T_{\bar{r}-\tilde{t}+i-1}^{(i-1)}} = \sigma_{\bar{r}-\tilde{t}+i}(Z) \quad \text{and} \quad \sum_{j=1}^{i-1} T_{\bar{r}-\tilde{t}+j}^{(i-1)} < \tilde{t} - \bar{r} + r.$$

Case 1: Assume $T_{\bar{r}-\tilde{t}+i}^{(i)}$ is such that $\sum_{j=1}^i T_{\bar{r}-\tilde{t}+j}^{(i)} = \tilde{t} - \bar{r} + r$ and

$$\sigma_{\bar{r}-\tilde{t}+i+1}(Z) < \frac{\sigma_{\bar{r}-\tilde{t}+i}(Z)}{1 + \gamma T_{\bar{r}-\tilde{t}+i}^{(i)}} = \frac{\sigma_{\bar{r}-\tilde{t}+j}(Z)}{(1 + \gamma T_{\bar{r}-\tilde{t}+j}^{(i-1)}) (1 + \gamma T_{\bar{r}-\tilde{t}+i}^{(i)})} = \frac{\sigma_{\bar{r}-\tilde{t}+j}(Z)}{1 + \gamma T_{\bar{r}-\tilde{t}+j}^{(i)}}, \quad 1 \leq j \leq i - 1.$$

Then, $i = \tilde{s}_{\tilde{t}}$ and $T_{\bar{r}-\tilde{t}+j} = T_{\bar{r}-\tilde{t}+j}^{(i)}$, $1 \leq j \leq i$, where

$$\begin{aligned} (1 + \gamma T_{\bar{r}-\tilde{t}+i-1}^{(i-1)}) (1 + \gamma T_{\bar{r}-\tilde{t}+i}^{(i)}) &= 1 + \gamma T_{\bar{r}-\tilde{t}+i-1}^{(i)} = 1 + \gamma \left(\tilde{t} - \bar{r} + r - T_{\bar{r}-\tilde{t}+i}^{(i)} - \sum_{j=1}^{i-2} T_{\bar{r}-\tilde{t}+j}^{(i)} \right) \\ &= i - 1 + \gamma \left(\tilde{t} - \bar{r} + r - T_{\bar{r}-\tilde{t}+i}^{(i)} \right) - \left(1 + \gamma T_{\bar{r}-\tilde{t}+i}^{(i)} \right) \sum_{j=1}^{i-2} \left(1 + \gamma T_{\bar{r}-\tilde{t}+j}^{(i-1)} \right) \end{aligned}$$

yields that

$$T_{\bar{r}-\tilde{t}+i}^{(i)} = \frac{\tilde{t} - \bar{r} + r - \sum_{j=1}^{i-1} T_{\bar{r}-\tilde{t}+j}^{(i-1)}}{i + \gamma \sum_{j=1}^{i-1} T_{\bar{r}-\tilde{t}+j}^{(i-1)}}$$

Case 2: Assume that there exists $T_{\bar{r}-\tilde{r}+i}^{(i)}$ such that for all j with $1 \leq j \leq i-1$

$$\sigma_{\bar{r}-\tilde{r}+i+1}(Z) = \frac{\sigma_{\bar{r}-\tilde{r}+i}(Z)}{1 + \gamma T_{\bar{r}-\tilde{r}+i}^{(i)}} = \frac{\sigma_{\bar{r}-\tilde{r}+j}(Z)}{\left(1 + \gamma T_{\bar{r}-\tilde{r}+j}^{(i-1)}\right) \left(1 + \gamma T_{\bar{r}-\tilde{r}+i}^{(i)}\right)} = \frac{\sigma_{\bar{r}-\tilde{r}+j}(Z)}{1 + \gamma T_{\bar{r}-\tilde{r}+j}^{(i)}}.$$

Then $i < \tilde{s}_{\tilde{r}}$ and we can set

$$T_{\bar{r}-\tilde{r}+i}^{(i)} = \gamma^{-1} \left(\frac{\sigma_{\bar{r}-\tilde{r}+i}(Z)}{\sigma_{\bar{r}-\tilde{r}+i+1}(Z)} - 1 \right).$$

In both cases

$$T_{\bar{r}-\tilde{r}+j}^{(i)} = \gamma^{-1} \left((1 + \gamma T_{\bar{r}-\tilde{r}+j}^{(i-1)})(1 + \gamma T_{\bar{r}-\tilde{r}+i}^{(i)}) - 1 \right), \quad 1 \leq j \leq i-1.$$

Eventually this procedure will find t, s and T that satisfy (44) – (46). Finally observe that

$$(47) \quad \frac{\sigma_{\bar{r}}(Z)}{1 + \gamma} > \sigma_{\bar{r}+1}(Z) \Rightarrow s = 0$$

in which case $\text{rank}(X^*) = \bar{r}$ and only t has to be determined. If additionally $\bar{r} = r$ then T is the identity matrix and finding t is redundant.

Algorithm 1 Determine $X = \text{prox}_{\frac{\gamma}{2}\|\cdot\|_F^2}(Z)$

```

1: Input: Let  $\gamma, r > 0$  and  $Z \in \mathbb{R}^{n \times m}$  be given and set  $\bar{r} = \lceil r \rceil$  and  $s = t = T_{\bar{r}} = 0$ .
2: Let  $Z = \sum_{i=1}^{\min\{m,n\}} \sigma_i(Z) u_i v_i^T$  be an SVD of  $Z$ .
3: while  $(\bar{r} > t \text{ AND } \sigma_{\bar{r}-t}(Z) \leq \frac{(1+\gamma)\sigma_{\bar{r}-t+1}(Z)}{1+\gamma T_{\bar{r}-t+1}})$  or  $t = 0$  do
4:    $T_{\bar{r}-t} = 0$ 
5:    $t = t + 1$ 
6:    $k = \bar{r} - t$ 
7:   while  $s \neq k$  do
8:      $k = k + 1$ 
9:      $T_k = \frac{t - \bar{r} + r - \sum_{j=\bar{r}+1-t}^{k-1} T_j}{t + k - \bar{r} + \gamma \sum_{j=\bar{r}+1-t}^{k-1} T_j}$ 
10:    if  $\frac{\sigma_k(Z)}{1 + \gamma T_k} \geq \sigma_{k+1}(Z)$  then
11:       $s = k$ 
12:    else
13:       $T_k = \gamma^{-1} \left( \frac{\sigma_k(Z)}{\sigma_{k+1}(Z)} - 1 \right)$ 
14:    end if
15:     $i = \bar{r} + t - 1$ 
16:    while  $i < k$  do
17:       $T_i = \gamma^{-1} ((1 + \gamma T_i)(1 + \gamma T_k) - 1)$ 
18:       $i = i + 1$ 
19:    end while
20:  end while
21: end while
22: Output:  $X = \frac{1}{1+\gamma} \sum_{i=1}^{\bar{r}-t} \sigma_i(Z) u_i v_i^T + \frac{\sigma_{\bar{r}}(Z)}{1+\gamma T_{\bar{r}}} \sum_{i=\bar{r}-t+1}^{\bar{r}+s} u_i v_i^T + \sum_{i=\bar{r}+s+1}^{\min\{m,n\}} \sigma_i(Z) u_i v_i^T$ .
```

Acknowledgments. We thank Andrey Ghulchak for his useful comments and numerous counter-examples. All authors are members of the LCCC Linnaeus Center and the eL-LIIT Excellence Center at Lund University. The third author is financially supported by the Swedish Foundation for Strategic Research.

REFERENCES

- [1] P.-A. ABSIL, R. MAHONY, AND R. SEPULCHRE, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2009.
- [2] A. C. ANTOUNAS, *On the approximation of Hankel matrices*, in *Operators, Systems and Linear Algebra*, Springer, 1997, pp. 17–22.
- [3] A. C. ANTOUNAS, *Approximation of large-scale dynamical systems*, vol. 6, SIAM, 2005.
- [4] A. ARGYRIOU, R. FOYCEL, AND N. SREBRO, *Sparse prediction with the k -support norm*, in *Advances in Neural Information Processing Systems*, 2012, pp. 1457–1465.
- [5] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer Science & Business Media, 2011.
- [6] C. BERGE, *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and convexity*, Courier Corporation, 1963.
- [7] M. W. BERRY, M. BROWNE, A. N. LANGVILLE, V. P. PAUCA, AND R. J. PLEMMONS, *Algorithms and applications for approximate nonnegative matrix factorization*, *Computational Statistics & Data Analysis*, 52 (2007), pp. 155–173.
- [8] S. BOYD, N. PARIKH, E. CHU, B. PELEATO, AND J. ECKSTEIN, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, *Foundations and Trends in Machine Learning*, 3 (2011), pp. 1–122.
- [9] S. P. BOYD AND L. VANDENBERGHE, *Convex optimization*, Cambridge University Press, 2004.
- [10] J.-F. CAI, E. J. CANDÈS, AND Z. SHEN, *A singular value thresholding algorithm for matrix completion*, *SIAM Journal on Optimization*, 20 (2010), pp. 1956–1982.
- [11] E. J. CANDÈS AND Y. PLAN, *Matrix completion with noise*, *Proceedings of the IEEE*, 98 (2010), pp. 925–936.
- [12] E. J. CANDÈS AND B. RECHT, *Exact matrix completion via convex optimization*, *Foundations of Computational mathematics*, 9 (2009), pp. 717–772.
- [13] E. J. CANDÈS AND T. TAO, *The power of convex relaxation: Near-optimal matrix completion*, *Information Theory, IEEE Transactions on*, 56 (2010), pp. 2053–2080.
- [14] V. CHANDRASEKARAN, S. SANGHAVI, P. A. PARRILO, AND A. S. WILLSKY, *Rank-sparsity incoherence for matrix decomposition*, *SIAM Journal on Optimization*, 21 (2011), pp. 572–596.
- [15] Y. CHEN, M. R. JOVANOVIĆ, AND T. T. GEORGIU, *State covariances and the matrix completion problem*, in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, IEEE, 2013, pp. 1702–1707.
- [16] M. T. CHU, R. E. FUNDERLIC, AND R. J. PLEMMONS, *Structured low rank approximation*, *Linear algebra and its applications*, 366 (2003), pp. 157–172.
- [17] P. L. COMBETTES AND J.-C. PESQUET, *Proximal splitting methods in signal processing*, in *Fixed-point algorithms for inverse problems in science and engineering*, Springer, 2011, pp. 185–212.
- [18] X. V. DOAN AND S. VAVASIS, *Finding the largest low-rank clusters with Ky Fan 2- k -norm and ℓ_1 -norm*, *SIAM J. Optim.*, 26 (2016), pp. 274–312.
- [19] J. DOUGLAS AND H. H. RACHFORD, *On the numerical solution of heat conduction problems in two and three space variables*, *Transactions of the American mathematical Society*, 82 (1956), pp. 421–439.
- [20] J. ECKSTEIN AND D. P. BERTSEKAS, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, *Mathematical Programming*, 55 (1992), pp. 293–318.
- [21] L. ELDEÉN, *Matrix methods in data mining and pattern recognition*, vol. 4, SIAM, 2007.
- [22] A. ERIKSSON, T. THANH PHAM, T.-J. CHIN, AND I. REID, *The k -support Norm and Convex Envelopes of Cardinality and Rank*, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3349–3357.
- [23] M. FAZEL, *Matrix rank minimization with applications*, PhD thesis, Stanford University, 2002.
- [24] M. FAZEL, H. HINDI, AND S. P. BOYD, *A rank minimization heuristic with application to minimum order system approximation*, in *American Control Conference, 2001. Proceedings of the 2001*, vol. 6, IEEE, 2001, pp. 4734–4739.
- [25] M. FREIMER AND G. S. MUDHOLKAR, *A class of generalizations of Hölder's inequality*, *Lecture Notes-Monograph Series*, (1984), pp. 59–67.
- [26] D. GABAY AND B. MERCIER, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, *Computers and Mathematics with Applications*, 2 (1976), pp. 17–40.
- [27] T. T. GEORGIU, *Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parametrization*, *IEEE transactions on Automatic Control*, 47 (2002), pp. 1811–1823.

- [28] T. T. GEORGIOU, *The structure of state covariances and its relation to the power spectrum of the input*, Automatic Control, IEEE Transactions on, 47 (2002), pp. 1056–1066.
- [29] R. GLOWINSKI AND A. MARROCO, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires*, ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique, 9 (1975), pp. 41–76.
- [30] C. GRUSSLER AND A. RANTZER, *Modified balanced truncation preserving ellipsoidal cone-invariance*, in 53rd IEEE Conference on Decision and Control, 2014, pp. 2365–2370.
- [31] C. GRUSSLER AND A. RANTZER, *On optimal low-rank approximation of non-negative matrices*, in Decision and Control (CDC), 2015 IEEE 54th Annual Conference on, IEEE, 2015, pp. 5278–5283.
- [32] C. GRUSSLER, A. ZARE, M. R. JOVANOVIĆ, AND A. RANTZER, *The use of the r^* heuristic in covariance completion problems*, in Submitted to Decision and Control (CDC), 2016 IEEE 55th Annual Conference on, IEEE, 2016, <http://www.ece.umn.edu/users/mihailo/papers/gruzarjovranCDC16.pdf>.
- [33] T. HASTIE, R. TIBSHIRANI, AND M. WAINWRIGHT, *Statistical learning with sparsity: the lasso and generalizations*, CRC Press, 2015.
- [34] R. HESSE AND D. R. LUKE, *Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems*, SIAM Journal on Optimization, 23 (2013), pp. 2397–2419.
- [35] R. HESSE, D. R. LUKE, AND P. NEUMANN, *Alternating projections and Douglas-Rachford for sparse affine feasibility*, Signal Processing, IEEE Transactions on, 62 (2014), pp. 4868–4881.
- [36] N. J. HIGHAM, *Computing the nearest correlation matrix – a problem from finance*, IMA journal of Numerical Analysis, 22 (2002), pp. 329–343.
- [37] J.-B. HIRIART-URRUTY AND C. LEMARCHAL, *Convex analysis and minimization algorithms. II. , Advanced theory and bundle methods*, Grundlehren der mathematischen Wissenschaften, Springer-Verlag, 1996.
- [38] J.-B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex analysis and minimization algorithms I: Fundamentals*, vol. 305, Springer Science & Business Media, 2013.
- [39] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, 2013.
- [40] A. J. IZENMAN, *Reduced-rank regression for the multivariate linear model*, Journal of multivariate analysis, 5 (1975), pp. 248–264.
- [41] J. KIM AND H. PARK, *Fast nonnegative matrix factorization: An active-set-like method and comparisons*, SIAM Journal on Scientific Computing, 33 (2011), pp. 3261–3281.
- [42] H. LAI, Y. PAN, C. LU, Y. TANG, AND S. YAN, *Efficient k -support matrix pursuit*, in Computer Vision–ECCV 2014, Springer, 2014, pp. 617–631.
- [43] V. LARSSON AND C. OLSSON, *Convex low rank approximation*, International Journal of Computer Vision, (2016), pp. 1–21.
- [44] V. LARSSON, C. OLSSON, E. BYLOW, AND F. KAHL, *Rank minimization with structured data patterns*, in Computer Vision–ECCV 2014, Springer, 2014, pp. 250–265.
- [45] F. LIN, M. R. JOVANOVIĆ, AND T. T. GEORGIOU, *An ADMM algorithm for matrix completion of partially known state covariances*, in Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on, IEEE, 2013, pp. 1684–1689.
- [46] P.-L. LIONS AND B. MERCIER, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis, 16 (1979), pp. 964–979.
- [47] X. LIU, Z. WEN, AND Y. ZHANG, *Limited memory block krylov subspace optimization for computing dominant singular value decompositions*, SIAM Journal on Scientific Computing, 35 (2013), pp. A1641–A1668.
- [48] D. G. LUENBERGER, *Optimization by vector space methods*, John Wiley & Sons, 1968.
- [49] I. MARKOVSKY, *Structured low-rank approximation and its applications*, Automatica, 44 (2008), pp. 891–909.
- [50] A. M. McDONALD, M. PONTIL, AND D. STAMOS, *New perspectives on k -support and cluster norms*, CoRR, abs/1512.08204 (2015), <http://arxiv.org/abs/1512.08204>.
- [51] B. MISHRA, G. MEYER, F. BACH, AND R. SEPULCHRE, *Low-rank optimization with trace norm penalty*, SIAM Journal on Optimization, 23 (2013), pp. 2124–2149.
- [52] C. OLSSON AND M. OSKARSSON, *A convex approach to low rank matrix approximation with missing data*, in Image Analysis, Springer, 2009, pp. 301–309.
- [53] J. R. PARTINGTON, *An Introduction to Hankel Operators*, vol. 13, Cambridge University Press, 1988.
- [54] D. PEACELLE, D. HENRION, Y. LABIT, AND K. TAITZ, *Users guide for SEDUMI interface 1.04*, LAAS-CNRS, Toulouse, (2002).
- [55] H. M. PHAN, *Linear convergence of the Douglas-Rachford method for two closed sets*. Available: <http://arxiv.org/abs/1401.6509>, Oct. 2014.
- [56] B. RECHT, M. FAZEL, AND P. A. PARRILO, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501.
- [57] R. T. ROCKAFELLAR, *Convex Analysis*, no. 28, Princeton University Press, 1970.
- [58] N. SREBRO, T. JAAKKOLA, ET AL., *Weighted low-rank approximations*, in ICML, vol. 3, 2003, pp. 720–727.
- [59] G. W. STEWART AND J.-G. SUN, *Matrix perturbation theory*, Academic press, 1990.

- [60] R. K. SUNDARAM, *A first course in optimization theory*, Cambridge university press, 1996.
- [61] R. TIBSHIRANI, *Regression shrinkage and selection via the Lasso*, Journal of the Royal Statistical Society. Series B (Methodological), (1996), pp. 267–288.
- [62] K.-C. TOH, M. J. TODD, AND R. H. TÜTÜNCÜ, *SDPT3 – a MATLAB software package for semidefinite programming, version 1.3*, Optimization methods and software, 11 (1999), pp. 545–581.
- [63] B. VANDEREYCKEN, *Low-rank matrix completion by riemannian optimization*, SIAM Journal on Optimization, 23 (2013), pp. 1214–1236.
- [64] R. VELU AND G. C. REINSEL, *Multivariate reduced-rank regression: theory and applications*, vol. 136, Springer Science & Business Media, 2013.
- [65] R. VIDAL, Y. MA, AND S. S. SASTRY, *Generalized Principal Component Analysis*, Springer, 2016.
- [66] G. A. WATSON, *Characterization of the subdifferential of some matrix norms*, Linear algebra and its applications, 170 (1992), pp. 33–45.
- [67] A. ZARE, Y. CHEN, M. R. JOVANOVIĆ, AND T. T. GEORGIU, *Low-complexity modeling of partially available second-order statistics: theory and an efficient matrix completion algorithm*, arXiv preprint arXiv:1412.3399, (2014).
- [68] A. ZARE, M. R. JOVANOVIĆ, AND T. T. GEORGIU, *Alternating direction optimization algorithms for covariance completion problems*, in American Control Conference (ACC), 2015, IEEE, 2015, pp. 515–520.
- [69] A. ZARE, M. R. JOVANOVIĆ, AND T. T. GEORGIU, *Color of turbulence*, arXiv preprint arXiv:1602.05105, (2016).
- [70] K. ZHOU, J. C. DOYLE, K. GLOVER, ET AL., *Robust and optimal control*, vol. 40, Prentice Hall, New Jersey, 1996.